

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:39:02 ; Search time 1878.33 Seconds
(without alignments)
-1517.972 Million cell updates/sec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcggacaaccccttgaaga.....aaatggcgatagatatattccc 2931

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_ba4.*
- 5: gb_ba5.*
- 6: gb_ba6.*
- 7: gb_ba7.*
- 8: gb_ba8.*
- 9: gb_ba9.*
- 10: gb_ba10.*
- 11: gb_ba11.*
- 12: gb_ba12.*
- 13: gb_ba13.*
- 14: gb_ba14.*
- 15: gb_ba15.*
- 16: gb_ba16.*
- 17: gb_ba17.*
- 18: gb_ba18.*
- 19: gb_ba19.*
- 20: gb_ba20.*
- 21: gb_ba21.*
- 22: gb_ba22.*
- 23: gb_ba23.*
- 24: gb_ba24.*
- 25: gb_ba25.*
- 26: gb_ba26.*
- 27: gb_ba27.*
- 28: gb_ba28.*
- 29: gb_ba29.*
- 30: gb_ba30.*
- 31: gb_ba31.*
- 32: gb_ba32.*
- 33: gb_ba33.*
- 34: gb_ba34.*
- 35: gb_ba35.*
- 36: gb_ba36.*
- 37: gb_ba37.*
- 38: gb_ba38.*
- 39: gb_ba39.*
- 40: gb_ba40.*
- 41: gb_ba41.*
- 42: gb_ba42.*
- 43: gb_ba43.*
- 44: gb_ba44.*

- 45: gb_ba45.*
- 46: gb_ba46.*
- 47: gb_ba47.*
- 48: gb_ba48.*
- 49: gb_ba49.*
- 50: gb_ba50.*
- 51: gb_ba51.*
- 52: gb_ba52.*
- 53: gb_ba53.*
- 54: gb_ba54.*
- 55: gb_ba55.*
- 56: gb_ba56.*
- 57: gb_ba57.*
- 58: gb_ba58.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2931	100.0	2931	8	AF053411	AF053411 Fucus dis
2	64.6	2.2	1794	5	E17199	E17199 Corallina p
3	64.6	2.2	2035	7	D87657	D87657 Corallina p
4	63.8	2.2	2029	7	D87658	D87658 Corallina p
5	62.8	2.1	1791	5	E17200	E17200 Corallina p
6	41.4	1.4	28043	42	AC014190	AC014190 Drosophila
7	41.4	1.4	75282	32	DMER37ML9	AL333495 Drosophila
8	38.6	1.3	38400	1	SC482	AL333495 Drosophila
9	38.6	1.3	38400	1	SC482	AL333495 Drosophila
10	38.6	1.3	38400	1	SC482	AL333495 Drosophila
11	38.6	1.3	38400	1	SC482	AL333495 Drosophila
12	38.6	1.3	38400	1	SC482	AL333495 Drosophila
13	38.6	1.3	38400	1	SC482	AL333495 Drosophila
14	38.6	1.3	38400	1	SC482	AL333495 Drosophila
15	38.6	1.3	38400	1	SC482	AL333495 Drosophila
16	38.6	1.3	38400	1	SC482	AL333495 Drosophila
17	38.6	1.3	38400	1	SC482	AL333495 Drosophila
18	38.6	1.3	38400	1	SC482	AL333495 Drosophila
19	38.6	1.3	38400	1	SC482	AL333495 Drosophila
20	38.6	1.3	38400	1	SC482	AL333495 Drosophila
21	38.6	1.3	38400	1	SC482	AL333495 Drosophila
22	38.6	1.3	38400	1	SC482	AL333495 Drosophila
23	38.6	1.3	38400	1	SC482	AL333495 Drosophila
24	38.6	1.3	38400	1	SC482	AL333495 Drosophila
25	38.6	1.3	38400	1	SC482	AL333495 Drosophila
26	38.6	1.3	38400	1	SC482	AL333495 Drosophila
27	38.6	1.3	38400	1	SC482	AL333495 Drosophila
28	38.6	1.3	38400	1	SC482	AL333495 Drosophila
29	38.6	1.3	38400	1	SC482	AL333495 Drosophila
30	38.6	1.3	38400	1	SC482	AL333495 Drosophila
31	38.6	1.3	38400	1	SC482	AL333495 Drosophila
32	38.6	1.3	38400	1	SC482	AL333495 Drosophila
33	38.6	1.3	38400	1	SC482	AL333495 Drosophila
34	38.6	1.3	38400	1	SC482	AL333495 Drosophila
35	38.6	1.3	38400	1	SC482	AL333495 Drosophila
36	38.6	1.3	38400	1	SC482	AL333495 Drosophila
37	38.6	1.3	38400	1	SC482	AL333495 Drosophila
38	38.6	1.3	38400	1	SC482	AL333495 Drosophila
39	38.6	1.3	38400	1	SC482	AL333495 Drosophila
40	38.6	1.3	38400	1	SC482	AL333495 Drosophila
41	38.6	1.3	38400	1	SC482	AL333495 Drosophila
42	38.6	1.3	38400	1	SC482	AL333495 Drosophila
43	38.6	1.3	38400	1	SC482	AL333495 Drosophila
44	38.6	1.3	38400	1	SC482	AL333495 Drosophila
45	38.6	1.3	38400	1	SC482	AL333495 Drosophila

ALIGNMENTS

```

RESULT 1
AF053411 2931 bp mRNA PLN 12-SEP-1998
LOCUS Fucus distichus vanadium bromoperoxidase mRNA, complete cds.
DEFINITION AF053411
ACCESSION AF053411
VERSION AF053411.1 GI:3582762
KEYWORDS
SOURCE
ORGANISM
Fucus distichus.
Fucus distichus
Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
Phaeophyceae; Fucales; Fucaceae; Fucus.
REFERENCE
1 (bases 1 to 2931)
AUTHORS Vreeland, V., Ng, K. and Epstein, L.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1998) Environmental Science, Policy and
Management, University of California, 201 Wellman Hall, Berkeley,
CA 94720-3112, USA
FEATURES
Location/Qualifiers
1..2931
/organism="Fucus distichus"
/db_xref="taxon:3012"
/dev_stage="2-cell embryo"
1..227
228..2258
/codon_start=1
/product="vanadium bromoperoxidase"
/protein_id="AAC35279.1"
/db_xref="GI:3582763"
/translaton="MLCHAAADTRGSPMDPTGVRLILTSORAKWRROLEKSLGF
HPSPTPIKYLEGSETHKWKVLPDTGISAKILGKIMARVRIATALAVLAAPCLAFD
EWTASGVFPEKHGTGEGHLOTCTNSDDALDPTAPNRDNVAFSRDRAARRERDGT
GTVCQITNGEDTLATMFHKSPLHDELGOVTDADFAILEDCLINGDFSICEVDYPPAGDA
GRLVNPAFAFDISGPAFASTIPPTLSPELAALAEALYMAALADYPPFMYGT
FTITTAANLACMGFPNLDVSIIGSDGVDPFVSOLFPAFTVGVGTGFPVOLLVNS
DEIDATVEKQETFAFDLNTWVDFEWNLONGPPGPEELDELFERFNARDJAR
VQFVNTINTEYRGSLLLELGAELGSRPCINGPFIDSDROAGFVNGTSHVFLIGAE
LAORASCYKOWHVRFAEPALGSLHTNTIAGLDLADFDJLLENDLLEKVAEINAA
QPNNEVYLLRQIJOVSPHPSYPSGHATONGAFATVLAJLIGLDGEGCFPNVP
PSDGLILNFEACLTYEGLNLAANVARGKGLGIHRYRFGIGLGLLGETIIVRT
LHQLMFAFEATFEFLFTGIVKILFVDGTFISIDGDMCSGLVITGVADCCQA
2259..2931 784 g 677 t
3'UTR 694 a 776 c
BASE COUNT
ORIGIN
Query Match 100.0%; Score 2931; DB 8; Length 2931;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cgcggacagccttggaagaggtgtgcccacattcaacagagcaggcccgaggtgt 60
Db 1 CGCGGACAGCCTTGGAAGAGAGGTTCGCCAATTCACAGAGCGAGGCCGCGTGAAGTGT 60
Qy 61 ggaagacagctgtacagctgtatccacagagcctcaactccctactatgacggagt 120
Db 61 GGAGGACAGCTGTCTACAGCTGATCCAGAGAGCGCTCAACTCCCTACTGATACGGAGT 120
Qy 121 ttatctgcgcgcgttcccaaaacagcaactttaacagcgtcggagcgcaatg 180
Db 121 TTGTACTGCGCGCGTTCGCAAAACCGCACTTTTAAACAGCGCTTCGCGAGCCACATG 180
Qy 181 ctccacagcatcccaaaaatcgacagtggtatcgctgagctggaatgctttgcattg 240
Db 181 CTTCACAGCATCCCAAAATCGACAGTGTATCGCTGAGCTGTAATGCTTTGGCCATG 240
Qy 241 cagcgacacagacagagcgctctctatgctgacacccgaggtgttcgttgctcaat 300
Db 241 CAGCGACACGACAGAGCGCTCTCTATGCTGTACACCGAGGTGCTCGTGTCTCAT 300
Qy 301 cagagcagcgcttaaggttggagcgcagttagaggggagagaatacactaggttttc 360
Db 301 CAGAGCAGCGCGCTAAAGTTGGAGCGCCAGTGTAGAGGGGAGAAATCACTAGGTTTC 360

```

```

Qy 361 atccaagcagagcgccttatataagctacttggaaggtctctgagacttgggaaggtta 420
Db 361 ATCCAAGCGAGAGCGCCTTATATCAAGTACTTGAAGGCTCTGAGACTTGAAGAGGTTA 420
Qy 421 agcttccaaacgagcgcgcatacgtctccaagatcctgggttaaaattatggcaggttcc 480
Db 421 AGCTTCCAACGAGCGGCATATCGGCTTCCAAAGTCTCTGGTAAATATATGGCAGGTC 480
Qy 481 gcatcgctacgcgccttggtgtggtactggcgccgacccctgttggcattcgacaggttca 540
Db 481 GCATCGCTACCGCTTGCTGGTGTGCTGCTGCTGCGCGCACCTGTCTTGGCATTCGACAGGTCA 540
Qy 541 cagcagctgtgttttccctgagacacacacacacacacacacacacacacacacacacacac 600
Db 541 CAGCAGCTGTGTGTTTCCCTGAGGAACACACACACACACACACACACACACACACACACAC 600
Qy 601 ccttataaaactcgcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 660
Db 601 CCTGTACAACTCCGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Qy 661 ctttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 661 CTTTGTGCTGCGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Qy 721 aaatcactaaacgagagaaactgattggctacacattgttccacagctctgcccacacagtg 780
Db 721 AAATCACTAAACGAGAGAACTGATTGGCTACATGTTCACAAAGTCTCTGTGCCACACGATG 780
Qy 781 aactgggacaggttaacgcgacgacttgcctatcctcagagactgcatcttaaacgag 840
Db 781 AACTGGGACAGGTAAACGCGACGACTTCGCTATCTCTCGAGGAGTGCATCTTTAAACGGAG 840
Qy 841 atttcagacatttcgagagcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 841 ATTTTCAGATTTCGAGAGAGCTGCTGCGGGAGAGCGCGCGGTGCGCTCGTCAATCTTA 900
Qy 901 cgcgcgttctccatccacatatacgcctccgcctcctcgcctcctcgcctcctcgcctcctcgc 960
Db 901 CGCGTGTGTTTGCATTCGACATATCCGGTCCCGATTCGCGCTAGGACAAATACCCCGG 960
Qy 961 tacctaccttctcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
Db 961 TACCTACCTTCTCTCTGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Qy 1021 tggccagggatgtaccctttatgcagtatggcaccgacgaattaccactaccgagcgag 1080
Db 1021 TGGCCAGGGATGTACCCTTTATGTCAGTATGGCACCCGAGAAATTTACCCTACCGCGGAG 1080
Qy 1081 ccaacccctgcctggaatggagcgttcccaaatctggagcgcgtgtcgataggggtccgatg 1140
Db 1081 CCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGACGCGCTGTCGATAGGGTCCGATG 1140
Qy 1141 gtacggtggacccgtctcctcagctcttcgcgagcacttcgttgggtgtgaaacggggc 1200
Db 1141 GTACGSGTGACCGCTTCTCCAGCTTTCGCGAGCGACTTCGTTGGTGTGTGAACGGGGC 1200
Qy 1201 ccttctcctcagctgctgtaaacagcttcaccatccacgcttaccgcttaccgcttaccgct 1260
Db 1201 CTTTCTCTCTCAGCTGCTGTAACAGCTTCACCATTCGACGCTATAGGCTGCAACCGA 1260
Qy 1261 agcaggaacattgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320
Db 1261 AGCAGGAGACATTCGCGCGCGCATTTGAACTATATGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Qy 1321 ttcgaagtgtggaaccccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1380
Db 1321 TTCGAATGTGTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Qy 1381 gtaacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440
Db 1381 GTAACGCCCGCGACCTTGGCGCAGGGTCTCTCTGCTGGGAAATATCAACACCGAAGCTTATC 1440

```


/protein_id="BAA31262.1"
/db_xref="GI:3293052"
/translation="MGIPADNLSQRAKASFTDTRVAAELALRGVVPFANGELLAR
NCTGDPSTIASFTKGLPDDNGAIIDPDDFLAFVPAIVNSGDEKEIADLTGLPARDE
TGLPIWRSDLANSLELVGWSNAGLTFTDLSGPAQVAMPAPVLMSPILIAEMA
ELYLALGRDIESEFSPKNOAIFRSALIRLNGLEWNTAKLGPDAIRERCEV
TVNLRGILPSEVGPVLYIIVGSKOIGSATVGNKTFVSPNADREFGEIAYSI
TISQVRIATPGDFWTLVFLDQADPRFSEYERGLINTIROLATVHERA
LPAZLANCLILLANGVPDNLPOQEDKLNDHDFVNFSAHVLSITVAVRAUK
AVQKFNHRLRERPEATGGLISVKNKSEPLAGSDILFPEVSELYELLSLIDVVAEN
EKORADGLVSPDKSELLPMAPAGSPFHSYSGHAYVAGACVILKAFDFANFQID
KVEVDTEKLVKSSFTGLTIVAGELNADNVAIGRNAGVHYFSDFESLLELQ
IAIGILEESGLIGENFFNLPFDGTTIQT"
449 a 524 c 349 g 507 t

Query Match 2.28; Score 63.8; DB 7; Length 2029;
Best Local Similarity 46.88; Pred. No. 1.4e-06;
Matches 240; Conservative 0; Mismatches 267; Indels 6; Gaps 1;
QY 1666 tcgcgggggagctagatgcagacttcgacatctccctcttctgaaatgatgagctctga 1725
DB 1404 TCCCTGAAGTGTAGTGAAGTGTGCGAGGAGCTCTCATCAATTCCTGGATGAGCTGTGCGAGA 1463
QY 1726 aacgtgtgagagataaattgcggcgcagataatcccaaacagagaggtcacctactcttc 1785
DB 1464 GCAATGAAACACAGACAGGCTGACGGGATTGTGAGCCCGATATAATCATTTCTCTTGC 1523
QY 1786 caaagctatcaagtgaggatcgcaacgacccctctaccctccgcgcacgctaccc 1845
DB 1524 CGATGCAATTTGCGAGGCGAGGCCATTCATCCGCTCGTATGGAAGTGGCGACGCTGG 1583
QY 1846 aaaaagagcatttcgacagcttcgaagccctctaccctccgcgcacgctaccc 1905
DB 1584 TTGCTGGCGCATGTGACAACTCTCAAGCGCTTTTCGACGCCA -----CTTCCAGA 1637
QY 1906 gttcccttaacccgtgtccaaagcgatgcagggctggaactaatcaacttcgaaggg 1965
DB 1638 TCGATAAGGTGTTGCGAGTGCACATCTCAAGCGCTTTTCGACGCCA -----CTTCCAGA 1637
QY 1966 catgcttaccatagaggagagatcaaacagctgcggtgcacagctgcacatttcggagggc 2025
DB 1698 GAACCTCTACATGTTGCGCGGTGAATTAACAAGCTCGCGACAAATGTTGCGATCGGCGGA 1757
QY 2026 agatgctgggcatccactatcggttcgacgggtatccaaagggcctactctcggagagacaa 2085
DB 1758 ACATGGCGGTGTTCACTACTTCTCTGACCACTTCGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1817
QY 2086 tcactgtacgaacttcaccagagagctgagcttcgagagagagagagagagagagagagagag 2145
DB 1818 TTGCGATGGAATCTTGGAGGACGAGCTGACGTATGCGGAGAACTTCTTCTTCAAT 1877
QY 2146 tcgcttattcacggagaggttcataacttt 2178
DB 1878 TCCCGAATTCAGGAAGTACATACATTCAGATTT 1910

RESULT 5
E17200 1791 bp DNA PAT 28-JUL-1999
LOCUS Corallina pilulifera mRNA for haloperoxidase.
DEFINITION E17200
ACCESSION E17200
VERSION E17200.1 GI:5711883
KEYWORDS JP 1998248581-A/2.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1791)
AUTHORS Izumi.Y. and Tanabe.T.
TITLE NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL Patent: JP 1998248581-A 22-SEP-1998;
OTSUKA PHARMACEUT CO LTD

COMMENT OS Corallina pilulifera
PN JP 1998248581-A/2
PD 22-SEP-1998
PF 06-MAR-1997 JP 1997070539
PI IZUMI YOSHIKAZU, TANABE TADASHI
PC C12N15/08,C12N9/08,(C12N15/08,C12R1:89),(C12N9/08,C12R1:19);
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FH source 1..1791
FH /clone="BPO2" /organism="Corallina pilulifera" FT
FH mat_peptide 1..1791
FH Location/Qualifiers
FH /product="haloperoxidase"
FH /organism="unidentified"
FH /db_xref="taxon:32644"
BASE COUNT 390 a 459 c 507 g 435 t
ORIGIN
Query Match 2.18; Score 62.8; DB 5; Length 1791;
Best Local Similarity 46.78; Pred. No. 2.7e-06;
Matches 239; Conservative 0; Mismatches 267; Indels 6; Gaps 1;
QY 1666 tcgcgggggagctagatgcagacttcgacatctccctcttctgaaatgatgagctctga 1725
DB 1286 TCCCTGAAGTGTAGTGAAGTGTGCGAGGAGCTCTCATCAATTCCTGGATGAGCTGTGCGAGA 1345
QY 1726 aacgtgtgagagataaattgcggcgcagataatcccaaacagagaggtcacctactcttc 1785
DB 1346 GCATGAAACACAGACAGGCTGACGGGATTGTGAGCCCGATATAATCATTTCTTGTGC 1405
QY 1786 caaagctatcaagtgaggatcgcaacgacccctctaccctccgcgcacgctaccc 1845
DB 1406 CGATGCAATTTGCGAGGCGAGGCCATTCATCCGCTCGTATGGAAGTGGCGACGCTGG 1465
QY 1846 aaaaagagcatttcgacagcttcgaagccctctaccctccgcgcacgctaccc 1905
DB 1466 TTGCTGGCGCATGTGACAACTCTCAAGCGCTTTTCGACGCCA -----CTTCCAGA 1519
QY 1906 gttcccttaacccgtgtccaaagcgatgcagggctggaactaatcaacttcgaaggg 1965
DB 1520 TCGATAAGGTGTTGCGAGTGCACACTGATGAGGACAAAGCTTGTGAAGTCTGTTCAAGG 1579
QY 1966 catgcttaccatagaggagagatcaaacagctgcggtgcacagctgcacatttcggagggc 2025
DB 1580 GAACCTCTACATGTTGCGCGGTGAATTAACAAGCTCGCGACAAATGTTGCGATCGGCGGA 1639
QY 2026 agatgctgggcatccactatcggttcgacgggtatccaaagggcctactctcggagagacaa 2085
DB 1640 ACATGGCGGTGTTCACTACTTCTCTGACCACTTCGAGTCTGAGTCTGAGTCTGAGTCTGAGT 1699
QY 2086 tcactgtacgaacttcaccagagagctgagcttcgagagagagagagagagagagagagagag 2145
DB 1700 TTGCGATGGAATCTTGGAGGAGGAGAGAGCTGACGTATGCGGAGAACTTCTTCTTCAAT 1759
QY 2146 tcgcttattcacggagaggttcataacttt 2177
DB 1760 TCCCGAATTCAGGAAGTACATACATTCAGATTT 1791
RESULT 6
AC014190 28043 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC014190
VERSION AC014190.1 GI:6437145
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 4H2 lies on the AseI B genomic fragment between St4H8 and St5B8.

FEATURES

source

1. 38400

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid 4H2"

1. 111

/note="overlap with Cosmid SC4H8 from: 15450 to: 15560"

complement(91..1086)

/gene="SC4H2.01c"

complement(91..1086)

/gene="SC4H2.01c"

/note="SC4H2.01c, probable secreted protein, len: 331 aa; some similarity to a secreted protein from Brucella abortus BOSP_BRUAB P12920 31 kd immunogenic protein precursor (329 aa), fasta scores; opt: 399 z-score: 357.4 E(): 1.1e-12, 32.7% identity in 336 aa overlap. Contains probable N-terminal signal sequence"

/codon_start=1

/transl_table=11

/product="putative secreted protein"

/protein_id="CAA18317.1"

/db_xref="GI:2995288"

/db_xref="SPTREMBL:O69956"

/translation="MSTLLSRIGRRVAVGVVGLLLWLLPLGKPPSGTIT FSTGRVYOKYGLLELTKRMDEGLDVLRLTMSAGSENVKVAAGDFTIAAD AVETLKLKAGSGAGLGRVTLDDYVQLVPSDSIOSVRLGRKVAIGLPNGYVR LIATLLKAVGIDPETDITPRADIDTGPRLGDELDAFFWGGVPGDGLKEAESSA FRFVPIGDDLVTLKRSRGATRYRTNPNESATYVONGAVVPTMAVSNLLTRADY DPLREWLTRVLDSDRGIGARVHSAQLVDLRLTIDPLEHEGARRYRSVKP"

complement(1095..1098)

/note="possible RBS upstream of SC4H2.01c"

1136..1675

/gene="SC4H2.02"

1136..1675

/note="SC4H2.02"

/gene="SC4H2.02, unknown, len: 179 aa; Proline-rich in N-terminus"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.02"

/protein_id="CAA18318.1"

/db_xref="GI:2995288"

/db_xref="SPTREMBL:O69957"

/translation="MAVATVGMSTTPAESGPPSAGRTPPDQGRTPDQGTSPAD LYRAHLAFLGDASTTETPPALAHRCCELLAEAAVAVYVSGCLDELAELADV VYVAGTALVGIIDLDVAVIAEVRHSNMTLGPDRVSRADGRVLAGDHTIFDVSGV LRQGWTPRTPDAPAPGA"

complement(1771..3006)

/gene="SC4H2.03c"

CDS

complement(1771..3006)

/gene="SC4H2.03c"

/note="SC4H2.03c, possible transmembrane transport protein, len: 411 aa, similar to several hypothetical proteins e.g. M.tuberculosis TR:O05881 (EMBL:U95121) M120B11.11C (385 aa), fasta scores opt: 1226 z-score: 1612.0 E(): 0, 35.5% identity in 372 aa overlap, and to members of the KRPB/KRCF potassium transporters family e.g. KRPB_ECOLI P45522 glutathione-regulated potassium-efflux protein (601 aa), fastascores; opt: 327 z-score: 292.9 E(): 4.3e-09, 25.3% identity in 396 aa overlap"

/codon_start=1

/transl_table=11

/product="putative transmembrane transport protein"

/protein_id="CAA18319.1"

/db_xref="GI:2995290"

/db_xref="SPTREMBL:O69958"

/translation="MHSVLLIEFGSIILGLLGRFAGRYRLSPIYLLAGLAFGE GLLPGASEFVATGAEIGVILLMLLGLVATSLDVLNKLKSHYPAGLVDCALNAVPGAVALLGMGVPAAVAVLAGVTWISSGVIAKVLGDLVGNRNPVTLVSLVEDLA MAYLLPITVLVAGVAGLMTGSITLAVLAAGLVFVAVYGRLLSRFVSSDDPEKLL LVVLGTLILVAGVAGLQVSAAGLVAGLSGEVAEGHLLSPRLDLFAAVFVFE FGHTPPASIPVPLPALALATVATKINTGYAARRAGISVKGWRAGCALVARGE FSVIAGLAVSAGIEPSGLATAYVLLVVLGLTARTFTEPLAMRLPFRDNDPRS GLAEGTPRAAPTEASVGD"

complement(3011..3496)

/gene="SC4H2.04c"

complement(3011..3496)

/gene="SC4H2.04c"

/note="SC4H2.04c, unknown, len: 161 aa; similar to hypothetical proteins from several organisms e.g. M. tuberculosis TR:O05882 (EMBL:U95121) MTCY20B11.12C (160 aa), fasta scores; opt: 315 z-score: 557.4 E(): 8e-24, 40.0% identity in 150 aa overlap (also upstream of a putative transport protein)"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.04c"

/protein_id="CAA18320.1"

/db_xref="GI:2995291"

/db_xref="SPTREMBL:O69959"

/translation="MSAPRLRATPLPGIYQVYDLVTRHRLSVVAHRDGARTVNVYR SDPDSCAHSLRLTGPEAGALIDALKPSHSHSLVTTDLGLVAERIEVAATSWNGR LLGDLRLTDTGASIVAVLRRAEIPSPAPDFRLVGGDLVIGTRGVDAARAILQG E"

complement(3014..3018)

gene

complement(3011..3496)

/gene="SC4H2.04c"

/note="SC4H2.04c, unknown, len: 161 aa; similar to hypothetical proteins from several organisms e.g. M. tuberculosis TR:O05882 (EMBL:U95121) MTCY20B11.12C (160 aa), fasta scores; opt: 315 z-score: 557.4 E(): 8e-24, 40.0% identity in 150 aa overlap (also upstream of a putative transport protein)"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.04c"

/protein_id="CAA18320.1"

/db_xref="GI:2995291"

/db_xref="SPTREMBL:O69959"

/translation="MSAPRLRATPLPGIYQVYDLVTRHRLSVVAHRDGARTVNVYR SDPDSCAHSLRLTGPEAGALIDALKPSHSHSLVTTDLGLVAERIEVAATSWNGR LLGDLRLTDTGASIVAVLRRAEIPSPAPDFRLVGGDLVIGTRGVDAARAILQG E"

complement(3014..3018)

CDS

complement(3011..3496)

/gene="SC4H2.04c"

/note="SC4H2.04c, unknown, len: 161 aa; similar to hypothetical proteins from several organisms e.g. M. tuberculosis TR:O05882 (EMBL:U95121) MTCY20B11.12C (160 aa), fasta scores; opt: 315 z-score: 557.4 E(): 8e-24, 40.0% identity in 150 aa overlap (also upstream of a putative transport protein)"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.04c"

/protein_id="CAA18320.1"

/db_xref="GI:2995291"

/db_xref="SPTREMBL:O69959"

/translation="MSAPRLRATPLPGIYQVYDLVTRHRLSVVAHRDGARTVNVYR SDPDSCAHSLRLTGPEAGALIDALKPSHSHSLVTTDLGLVAERIEVAATSWNGR LLGDLRLTDTGASIVAVLRRAEIPSPAPDFRLVGGDLVIGTRGVDAARAILQG E"

complement(3014..3018)

RBS

complement(3014..3018)

/gene="SC4H2.04c"

/note="possible RBS upstream of SC4H2.03c"

complement(3506..3509)

/note="possible RBS upstream of SC4H2.04c"

3602..3606

RBS

complement(3506..3509)

/note="possible RBS upstream of SC4H2.05"

3611..4687

/gene="SC4H2.05"

3611..4687

/gene="SC4H2.05"

/note="SC4H2.05, probable two-component sensor protein, len: 358 aa; similar to many e.g. DEGS_BACSU P13799 sensor protein degs (ec 2.7.3.-), Bacillus subtilis (385 aa), fasta scores; opt: 317 z-score: 334.6 E(): 2e-11, 31.3% identity in 227 aa overlap"

/codon_start=1

/transl_table=11

/product="putative two-component sensor"

/protein_id="CAA18321.1"

/db_xref="GI:2995292"

/db_xref="SPTREMBL:O69960"

/translation="MGVYDTRSPRSKALACVSCKAGRSKENVSLFWRFSLNA AGLYVATALLGPVTPVPLVAGLVLLAGLAAAGLVAGNVVRLGTPVLRIGLRAMS TALLVPGTRPVSQPAQAQLIATVMTLMDLQASAGACAGLQAEPRHRIARE LHDEVGITLTVLQGLKRVADRVFGEIRLDEVTIQAQETAGLDEIRRLRRKRGVLE ELGASALSLANETFHGLTVQHIFGDFLPLAPAEVLVIRVAQEGELINTARHADA

gene

complement(1095..1098)

/note="possible RBS upstream of SC4H2.01c"

1136..1675

/gene="SC4H2.02"

1136..1675

/note="SC4H2.02"

/gene="SC4H2.02, unknown, len: 179 aa; Proline-rich in N-terminus"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.02"

/protein_id="CAA18318.1"

/db_xref="GI:2995288"

/db_xref="SPTREMBL:O69957"

/translation="MAVATVGMSTTPAESGPPSAGRTPPDQGRTPDQGTSPAD LYRAHLAFLGDASTTETPPALAHRCCELLAEAAVAVYVSGCLDELAELADV VYVAGTALVGIIDLDVAVIAEVRHSNMTLGPDRVSRADGRVLAGDHTIFDVSGV LRQGWTPRTPDAPAPGA"

complement(1771..3006)

/gene="SC4H2.03c"

CDS

complement(1771..3006)

/gene="SC4H2.03c"

/note="SC4H2.03c, possible transmembrane transport protein, len: 411 aa, similar to several hypothetical proteins e.g. M.tuberculosis TR:O05881 (EMBL:U95121) M120B11.11C (385 aa), fasta scores opt: 1226 z-score: 1612.0 E(): 0, 35.5% identity in 372 aa overlap, and to members of the KRPB/KRCF potassium transporters family e.g. KRPB_ECOLI P45522 glutathione-regulated potassium-efflux protein (601 aa), fastascores; opt: 327 z-score: 292.9 E(): 4.3e-09, 25.3% identity in 396 aa overlap"

/codon_start=1

/transl_table=11

/product="putative transmembrane transport protein"

/protein_id="CAA18319.1"

/db_xref="GI:2995290"

/db_xref="SPTREMBL:O69958"

/translation="MHSVLLIEFGSIILGLLGRFAGRYRLSPIYLLAGLAFGE GLLPGASEFVATGAEIGVILLMLLGLVATSLDVLNKLKSHYPAGLVDCALNAVPGAVALLGMGVPAAVAVLAGVTWISSGVIAKVLGDLVGNRNPVTLVSLVEDLA MAYLLPITVLVAGVAGLMTGSITLAVLAAGLVFVAVYGRLLSRFVSSDDPEKLL LVVLGTLILVAGVAGLQVSAAGLVAGLSGEVAEGHLLSPRLDLFAAVFVFE FGHTPPASIPVPLPALALATVATKINTGYAARRAGISVKGWRAGCALVARGE FSVIAGLAVSAGIEPSGLATAYVLLVVLGLTARTFTEPLAMRLPFRDNDPRS GLAEGTPRAAPTEASVGD"

complement(3011..3496)

/gene="SC4H2.04c"

complement(3011..3496)

/gene="SC4H2.04c"

/note="SC4H2.04c, unknown, len: 161 aa; similar to hypothetical proteins from several organisms e.g. M. tuberculosis TR:O05882 (EMBL:U95121) MTCY20B11.12C (160 aa), fasta scores; opt: 315 z-score: 557.4 E(): 8e-24, 40.0% identity in 150 aa overlap (also upstream of a putative transport protein)"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.04c"

/protein_id="CAA18320.1"

/db_xref="GI:2995291"

/db_xref="SPTREMBL:O69959"

/translation="MSAPRLRATPLPGIYQVYDLVTRHRLSVVAHRDGARTVNVYR SDPDSCAHSLRLTGPEAGALIDALKPSHSHSLVTTDLGLVAERIEVAATSWNGR LLGDLRLTDTGASIVAVLRRAEIPSPAPDFRLVGGDLVIGTRGVDAARAILQG E"

complement(3014..3018)

gene

complement(1771..3006)

/gene="SC4H2.03c"

/note="SC4H2.03c, possible transmembrane transport protein, len: 411 aa, similar to several hypothetical proteins e.g. M.tuberculosis TR:O05881 (EMBL:U95121) M120B11.11C (385 aa), fasta scores opt: 1226 z-score: 1612.0 E(): 0, 35.5% identity in 372 aa overlap, and to members of the KRPB/KRCF potassium transporters family e.g. KRPB_ECOLI P45522 glutathione-regulated potassium-efflux protein (601 aa), fastascores; opt: 327 z-score: 292.9 E(): 4.3e-09, 25.3% identity in 396 aa overlap"

/codon_start=1

/transl_table=11

/product="putative transmembrane transport protein"

/protein_id="CAA18319.1"

/db_xref="GI:2995290"

/db_xref="SPTREMBL:O69958"

/translation="MHSVLLIEFGSIILGLLGRFAGRYRLSPIYLLAGLAFGE GLLPGASEFVATGAEIGVILLMLLGLVATSLDVLNKLKSHYPAGLVDCALNAVPGAVALLGMGVPAAVAVLAGVTWISSGVIAKVLGDLVGNRNPVTLVSLVEDLA MAYLLPITVLVAGVAGLMTGSITLAVLAAGLVFVAVYGRLLSRFVSSDDPEKLL LVVLGTLILVAGVAGLQVSAAGLVAGLSGEVAEGHLLSPRLDLFAAVFVFE FGHTPPASIPVPLPALALATVATKINTGYAARRAGISVKGWRAGCALVARGE FSVIAGLAVSAGIEPSGLATAYVLLVVLGLTARTFTEPLAMRLPFRDNDPRS GLAEGTPRAAPTEASVGD"

complement(3011..3496)

/gene="SC4H2.04c"

complement(3011..3496)

/gene="SC4H2.04c"

/note="SC4H2.04c, unknown, len: 161 aa; similar to hypothetical proteins from several organisms e.g. M. tuberculosis TR:O05882 (EMBL:U95121) MTCY20B11.12C (160 aa), fasta scores; opt: 315 z-score: 557.4 E(): 8e-24, 40.0% identity in 150 aa overlap (also upstream of a putative transport protein)"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC4H2.04c"

/protein_id="CAA18320.1"

/db_xref="GI:2995291"

/db_xref="SPTREMBL:O69959"

/translation="MSAPRLRATPLPGIYQVYDLVTRHRLSVVAHRDGARTVNVYR SDPDSCAHSLRLTGPEAGALIDALKPSHSHSLVTTDLGLVAERIEVAATSWNGR LLGDLRLTDTGASIVAVLRRAEIPSPAPDFRLVGGDLVIGTRGVDAARAILQG E"

complement(3014..3018)

AAALTLHLPLGAGVELLVRDDTGLGTAAAGAGIRGMRERALLIGAETHPEAPGPG
GTDVRLRPVAPPGRCDADRTGDSF

gene

4687..5346

CDS

/gene="SC4H2.06"

/gene="SC4H2.06"

/note="SC4H2.06, probable two-component transcriptional
regulator, len: 219 aa; similar to many e.g. DREGU, BACSU
P13800 transcriptional regulatory protein degu (229 aa),
fasta scores: opt: 343 z-score: 736.1 E(1): 0, 37.18
identity in 224 aa overlap. Contains F50622 Bacterial
regulatory proteins, luxR family signature, Pfam match to

Query Match 1.3% Score 39.4; DB 1; Length 38400;
Best Local Similarity 44.2%; Pred. No. 12; Mismatches 206; Indels 0; Gaps 0;
Matches 163; Conservative 0;

QY 1150 accgtctccagctctccagcgagccttctgtgttgaaacggggccctttgtct 1209

DB 13012 ACCGCTCGCGGTCACCTGATCTCGCCGAACCTGGCGGAGACACACCCCTGACCG 13071

QY 1210 ctgagctgctgtaaacagcttccaccatcgacgctattacggtcgaaacgagcaggaga 1269

DB 13072 CCGTCTCTCCACGACACCGCTCGAAGACACCGAGGTGACGCTCGATCAGGTGGCGGAGC 13131

QY 1270 cattcgccccgactgaactatggtctgattttgacgaatggtgaacattcgaatg 1329

DB 13132 AGTTTCGCGCGGAGGTCCCTACCTCTGCGAGCGCTCACCAAGCTGGAGAGTGGACT 13191

QY 1330 gtgagccccgccccgccccgaggttagcagagagctgtttttatcgttaacagccc 1389

DB 13192 ACGCGCGCGCGGAGCGCGAGACCTTCGCAAGATGCTCTCGCCACCGGACGAGC 13251

QY 1390 gcaactggcagaggtctctctgtgacaatatcaacaggaagcttatcgaggtctc 1449

DB 13252 TCCGCGTGTGTCATCAACTTCGCGAGCGGCTGCACATCGCCACCTCTGTGTCA 13311

QY 1450 ttatctactgagctggagcttcacagcccggtatcaacggttcattcatcgaca 1509

DB 13312 TGCCTCCGAGAGAGGAGCAACCATCCCAAGGTACCCCGGACGCTCTCATCCGCTCG 13371

QY 1510 gtgatcgcc 1518

DB 13372 CCGAACGCGC 13380

RESULT 9

SCC88/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

SCC88 42816 bp DNA BCT 15-FEB-2000

Streptomyces coelicolor cosmid C88.

AL139298

AL139298.1 GI:6983731

50S ribosomal protein L21; 50S ribosomal protein L27;

dehydrogenase; fpgs; glycosyltransferase; integral membrane

phb2; secreted protein; sfr; transferase; vals.

Streptomyces coelicolor A3(2).

Streptomyces coelicolor A3(2).

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 42816)

Kinashi, H. and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map

for the 8 Mb Streptomyces coelicolor A3(2) chromosome

Mol. Microbiol. 21 (1), 77-96 (1996)

97000351

2 (bases 1 to 42816)

Saunders, D. and Harris, D.

Unpublished

3 (bases 1 to 42816)

Cardeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Direct Submission

JOURNAL

COMMENT

Submitted (14-FEB-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at the Sanger Centre is funded
by the BBSRC and Rosalind Franklin
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984), as implemented at
<http://www.nih.gov/jp/>

fun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon) if this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: this sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid C88 lies between and overlaps with cosmids C123 and C80 on
the AseI-C genomic restriction fragment.

Location/Qualifiers
1. 42816
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid C88"
1. 3409
/gene="SCC88.01"
1. 106
/note="nominal overlap with S. coelicolor cosmid StC123"
1. 3409
/partial
/gene="SCC88.01"
/note="Protein sequence is in conflict with the conceptual
translation: SCC88.01 possible glycosyltransferase, len:
>1135 aa; N-terminal region similar to TR:Q9VC92
(EMBL:AF146532) Klebsiella pneumoniae glycosyltransferase,
329 aa; fasta scores: opt: 214 z-score: 245.4 E(1): 3e-06;
21.3% identity in 263 aa overlap and C-terminal region
similar to SW:TAGF_BACSU (EMBL:X15200) Bacillus subtilis
tetraoic acid biosynthesis protein F TagF, 746 aa; fasta
scores: opt: 1056 z-score: 1203.0 E(1): 0; 41.5% identity
in 381 aa overlap"
/codon_start=2
/transl_table=11
/product="putative glycosyltransferase"
/protein_id="CAB75371.1"
/db_xref="GI:6983732"

FEATURES

source

gene

misc_feature

CDS

IQGRRSAKVLVKKDGSRRATVPLNTHVPEATSLGRKHNTDMSGNAADLP
 LKLRGSEWERTWVGIGIATAGLLRKAVOVKGPRAHPPVOMLDDDFLLPTPV
 NGALRLTAVRALITGHRAVAGIDELGRLVPLGAGESVSLRVTKNKHLPV
 TISAADGATPTTVVRVPMRDVALLPHODVYKEMAAGDPAARSRNWSLTOLVAV
 PDGGRSTVTVREGIADGEMRLPASLAENADREIALVSHGSHYKLFPSRLOKAIT
 SVRTGDDVYVGGSAADIDGADVVKARDVDFEKTVALRTSPDGFAAFETPASMGS
 PDGTLKSGWNEFLPTDGHIDPVLDRVADGFPVLGNAKDREYEPESWYFPO
 LKPSDLSELRGAYROERLRKDYVAGTPTOLRDVLFVSYNGKQYSDSPRAHEEL
 LRAGDLEVLWVDEGVDLPDPAOKPWPASTKYEALASCRVIVTNAHLPGWERRP
 GOVIVHCTGMLKIGLIDIEAPKEDROYHERLTOEARNTMLVSNRSTILKARRP
 GFDKLTGVPNDXLYAPDLAEARELKERLGVFSDKRVILHATWDDLSHARGQ
 FKRLRIDEVARARLGDDEVLRLRHSNYVDSIPGAGNFVHDVSEYTDIADLLIAA
 DVLVDTSSVNFDAHLRPMFLFTDLEHTDKRLGFIYDFENDAPGPIINTSEELY
 SALGDLKVSATYADRIIRRFOLFCDLDDGAAARVVDLMLEQAREV*
 3416..3460
 /gene="SCC88.02"
 /note="SCC88.02"
 /db_xref="GI:6983733"

stem_loop.
 gene
 CDS

/note="SCC88.02. possible secreted protein, len: 576 aa;
 similar to TR:CA51964 (EMBL:AL010961) Streptomyces
 coelicolor hypothetical 100.7 kD protein SC6E10.10, 943
 aa; fasta scores: opt: 522 z-score: 556.2 E(): 1.4e-23;
 31.8% identity in 471 aa overlap. Contains possible
 N-terminal region signal peptide sequence"

/codon_start=1
 /transl_table=11
 /product="putative secreted protein"
 /protein_id="CA575372.1"
 /db_xref="GI:6983733"

/translation="MFTSRARLRTRSLTCTALVAAGMLVTPPAADKPSPIE
 VTPKSQLPLDRAOPKQAPRMAATPAFREDVDGDSMLRLADGSGVGLT
 TADGFTPTLVNDVYERKQDVTIPGDLGNGPEILTLSTGTLSPGMSGADSTV
 ATWSTGWLQVNVKVPAGVDGTPDGLLARTYGGDLXYQGSVDSPPRGKSTVG
 GWEMDOLVGMNGDGLGVVATRLGLYLYFVAGIGSATPLANRDLGCTGYN
 QILAADWDGDLGLLGRTRSGDLYVRAAGTGNFPAKQAGTQPVYFAGSGVY
 PHEKASLVGMQTRGLYXVYKNGQLTGREQSDTGGSGRAYVATSLDQSIAD
 LLELYDVLVYKANTADESLAGKINFLFQPLDADGSLANDLDSGLHFLGR
 KHGASLSLKLKVGWGNAYDAGVDFGTGDRIDYVATPKAKLADYSGVLSIP
 PAGVLDGSGWQVKNLAFSGDIDGGRADLLAVNSGGDLIRITSTIDGRFAARVLG
 TGNVIRNLX
 5327..5362
 /complement(5387..8215)
 /note="SCC88.03c, possible transferase, len: 942 aa;
 N-terminal region similar to hypothetical proteins, e. g.
 TR:069853 (EMBL:AL023702) Streptomyces coelicolor
 hypothetical 66.0 kD protein SC1C3.11, 586 aa; fasta
 scores: opt: 1214 z-score: 1395.6 E(): 0; 40.7% identity
 in 573 aa overlap and C-terminal region similar to
 TR:069854 (EMBL:AL023702) Streptomyces coelicolor putative
 transferase SC1C3.12, 697 aa; fasta scores: opt: 658
 z-score: 754.0 E(): 0; 30.1% identity in 574 aa overlap.
 Contains pfam match at N-terminal region to entry PF00534
 Glycosyl transferase group 1"

stem_loop
 gene
 CDS

/codon_start=1
 /transl_table=11
 /product="putative transferase"
 /protein_id="CA575373.1"
 /db_xref="GI:6983734"

/translation="MKITFLLTWGDMGTQAVYVTOATHLAPRHIEVLVFTREQ
 PFFSVDERSVRLVDRTKTERVRESDLTAECRLAGLPSDLSPKWEATFDALS
 DYEMORALITDITLTPSPALSAVADLAPSRVITQQEHRPSOLRGTEPELLR
 APEADLVITERITKWLSESLGKAAPLAAIPNAIPEGFRPSLTGKITVMPRLV
 PDKQVRAIQAFAPLDPHWRILFDFGDPMSRLNLTQGLHDSVYELGSPQHM
 TEWARSLALFSPQGEAPFVILLEAFAGVAPAYDVIYTPAEIIRGDEGLGLV
 NMESARALSRMGDEALRLSYGEKHEGSTFAADVIYKWEELFVLSYRRDDR
 RFAERLARIHVAHGAGRFAAADAORTAPSSQDQREYVIGASRLSVRAGGLR
 SEVRLDQSGEIVQNFETVWEALSSGIPYVLLRDRDNNRRLAVDAAEQTRVKA
 LAGAYEKAVYALLPRTHAPGVLLAEVGAAGVAGVFPVYVSTPLRFGPA
 GCDIEFWQVPEEGDGGQFVAPRPSAVGPKPLSLTPDARTVAKDREYPTLELTRK

LVSDITFPDVAITYTVDSDPRWQERRARRAALGLEAESGDEAARNRDELRYSL
 RSLANFQSKIYLVTDOTPEMLNTEHEGLVVSHRDITDDOCLPTNSHSIESQ
 LHHDGLSEQLINDVDFTRGPQAGREFLNGASREFWSPPTVPGEETDEDEYF
 AKANRRLEREGYTVANSFVAPHLRSLVLSISEDPESVATATPWRQOD
 ISWASLHHHYGLTGAQVSSSIRKAYIDVGYSHRPELTALLARGHDFCLGESOD
 AVPHEQARIVAEFLAPRFPKSPYER*
 complement(7124..7570)
 /gene="SCC88.03c"
 /note="Pfam matches to entry PF00534 Glycosyl transferase group 1, score 99.70, E-value 2.1e-27"

misc_feature

/complement(8222..9211)
 /gene="SCC88.04c"
 /complement(8222..9211)
 /note="SCC88.04c, unknown, len: 329 aa"

gene
 CDS

/codon_start=1
 /transl_table=11
 /product="hypothetical protein SCC88.04c"
 /protein_id="CA575374.1"
 /db_xref="GI:6983735"

/translation="MKOLTPPHQVMRKYAARSRLDSLVLPSPDDWPRVAAAD
 HSLDGLTANVCTVPDDITAAASLVLGSGSPVPLTLVORGDRVEARGTSV
 DLLENVRADSPVPAGMRRLLEAGOWRLSVVTVDRGERRFALAAAPRLMPDGPTL
 PEPVADGTICRVASSTPGRAYLSLGRDHAAEIVSDIGSQITLGRVLNPSPAEC
 SGDELIRRNKSVKSTPLTATWQGDVFTCTVHTADFGASGSEQIWDVLRRLRPRSLK
 LCRRITDVRSPDVRPMPNLLTADDTALRINPYTTPVGSFLAFRAALIPSGS"
 complement(9211..11019)
 /gene="SCC88.05c"

gene

Query Match: 1.3% Score 38.6; DB 2; Length 42816;
 Best Local Similarity 51.4%; Pred No. 21;
 Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 745 tggctacatgtctccacaagtctctgcacacgatgaactggagcaggttaaccgagacg 804
 DB 9138 TGGACACCTCGCTCGTCCCGCGAGTCTCTGACACTGCGCCGCGCGCGGACC 9079
 QY 805 attctgctatctcggagcagtcacatttaacgagagatttcagcatttgcgagagcgtgc 864
 DB 9078 ACCTCTCCCTCCGAGCGCGCCGACCTCAACGCTCGCTCGGACCGCGCGGACCTCA 9019
 QY 865 ctgcgggagacccggcggtctcctcgtcaatcctaccgctgctgttgcctc 917
 DB 9018 CCGCGCGGCTCCGCGCTCGTCTCGCGCGGCTCTCCCGCGCTCGTGATC 8966

RESULT 10
 AC022255/c

LOCUS Homo sapiens chromosome 4 clone RP11-440L14 map 4, WORKING DRAFT
 DEFINITION AC022255 189975 bp DNA HTG 01-FEB-2000
 SEQUENCE 20 unordered pieces.

ACCESSION AC022255

VERSION AC022255.2 GI:6850411

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT

SOURCE human

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 189975)

2 (bases 1 to 189975)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Becker, R., Bada, F.,

Boguslavsky, L., Bouckhagalter, B., Brown, A., Burkett, G., Castle, A.,

Choe, Y., Colangelo, M., Collins, S., Collins, S., Collymore, A., Cooke, P.,

DeArnell, K., Dewar, K., Domino, M., Doyle, M., Fenster, J.,

Ferreira, P., Fitch, W., Forrest, C., Gage, D., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
Travers, M., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

TITLE
JOURNAL.

COMMENT

2002/11/14
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A. P. & Green, P. (1996-1997).

Smith, K.T., A. Green, & (1996-1997).
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6506
 Center clone name: 589_P_23

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved

[illegible]

19020	19944:	contig of 925 bp in length
*	*	gap of unknown length
19945	20863:	contig of 919 bp in length
*	*	gap of unknown length
20864	21773:	contig of 910 bp in length
*	*	gap of unknown length
21774	22686:	contig of 913 bp in length
*	*	gap of unknown length
22687	23617:	contig of 931 bp in length
*	*	gap of unknown length
23618	24528:	contig of 911 bp in length
*	*	gap of unknown length
24529	25425:	contig of 897 bp in length
*	*	gap of unknown length
25426	26351:	contig of 926 bp in length
*	*	gap of unknown length
26352	27273:	contig of 922 bp in length
*	*	gap of unknown length
27274	28210:	contig of 937 bp in length
*	*	gap of unknown length
28211	29139:	contig of 929 bp in length
*	*	gap of unknown length
29140	30047:	contig of 908 bp in length
*	*	gap of unknown length
30048	30908:	contig of 861 bp in length
*	*	gap of unknown length
30909	31830:	contig of 922 bp in length
*	*	gap of unknown length
31831	32733:	contig of 903 bp in length
*	*	gap of unknown length
32734	33629:	contig of 896 bp in length
*	*	gap of unknown length
33630	34512:	contig of 883 bp in length
*	*	gap of unknown length
34513	35438:	contig of 926 bp in length
*	*	gap of unknown length
35439	36308:	contig of 870 bp in length
*	*	gap of unknown length
36309	37230:	contig of 922 bp in length
*	*	gap of unknown length
37231	38157:	contig of 927 bp in length
*	*	gap of unknown length
38158	39069:	contig of 912 bp in length
*	*	gap of unknown length
39070	40024:	contig of 955 bp in length
*	*	gap of unknown length
40025	40913:	contig of 889 bp in length
*	*	gap of unknown length
40914	41801:	contig of 888 bp in length
*	*	gap of unknown length
41802	42710:	contig of 909 bp in length
*	*	gap of unknown length
42711	43661:	contig of 951 bp in length
*	*	gap of unknown length
43662	44541:	contig of 880 bp in length
*	*	gap of unknown length
44542	45435:	contig of 894 bp in length
*	*	gap of unknown length
45436	46349:	contig of 914 bp in length
*	*	gap of unknown length
46350	47269:	contig of 920 bp in length
*	*	gap of unknown length
47270	48204:	contig of 935 bp in length
*	*	gap of unknown length
48205	49085:	contig of 881 bp in length
*	*	gap of unknown length
49086	49990:	contig of 905 bp in length
*	*	gap of unknown length
49991	50894:	contig of 904 bp in length
*	*	gap of unknown length
50895	51781:	contig of 887 bp in length
*	*	gap of unknown length
51782	52679:	contig of 898 bp in length

```

* 52680 53552: gap of unknown length
* 53553 54466: contig of 873 bp in length
* 54467 55364: contig of 914 bp in length
* 55365 56295: contig of 898 bp in length
* 56296 57215: contig of 931 bp in length
* 57216 58155: contig of 920 bp in length
* 58156 59072: contig of 940 bp in length
* 59073 59968: contig of 917 bp in length
* 59969 60878: contig of 896 bp in length
* 60879 61785: contig of 910 bp in length
* 61786 62664: contig of 907 bp in length
* 62664: contig of 879 bp in length
* 62664: gap of unknown length

Query Match 1.38; Score 37.6; DB 55; Length 72683;
Best Local Similarity 52.68; Pred. No. 43;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 2774 tggagtagctgtaccagtaggtgtaagacatatcttctctgtatgctgtaga 2833
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43942 TGAAGAAATGGAAGACTTGGCTCTCTAATAATACATCTTCGATCTTCCAA 44001
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 2834 acagctagctgtttttatcacagatgctataaataaggatggtgataatggcgc 2893
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44002 GGAGCTATTGTTATTTTACCAAGATTTCCAGTAATAGAAATTCAAATATACTGT 44061
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 2894 ggtactatgaacgcgaataaaggcagatattc 2929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44062 GAGCTTCTAAACACCATATTCATCTGCTGATATAC 44097
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12

```

HS440021 107677 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 440021 on chromosome X contains ESTs
and STs.
ACCESSION 284481
VERSION 284481.1 GI:1945153
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 107677)
Pearce/A.
Direct Submission
Submitted (17-APR-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 19, 1997 this sequence version replaced gi:1806016.
de Jong P. J. enquiries: http://bacpac.med.buffalo.edu/IMPORTANT:
This sequence is the entire insert of clone 440021. This sequence
has been finished according to sequence map criteria as follows.
An attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 440021 is at 1 in this sequence. The
true right end of clone 440021 is at 107677.
440021 is from the library RPCI3 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pcypac2.

```

FEATURES source

```

Location/Qualifiers
1..107677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="RP3-440021"
/clone_lib="RPCI-3"
117..207
repeat_region
117..207
/note="L1M4 repeat: matches 961..1047 of consensus"
331..1332
repeat_region
331..1332
/note="L1 repeat: matches 4392..5390 of consensus"
1179..2209
repeat_region
1179..2209
/note="L1M3 repeat: matches 1..1051 of consensus"
2917..3215
repeat_region
2917..3215
/note="AluX repeat: matches 5..302 of consensus"
3668..3877
repeat_region
3668..3877
/note="MIR repeat: matches 247..41 of consensus"
4993..5216
repeat_region
4993..5216
/note="Alu repeat: matches 301..80 of consensus;
incomplete repeat"
5239..5279
repeat_region
5239..5279
/note="MLT2E repeat: matches 397..356 of consensus"
5343..5789
repeat_region
5343..5789
/note="MLT2FA repeat: matches 395..8 of consensus"
7550..7857
repeat_region
7550..7857
/note="Alu repeat: matches 295..1 of consensus"
9156..9324
repeat_region
9156..9324
/note="L1 repeat: matches 4871..5035 of consensus"
9652..9698
repeat_region
9652..9698
/note="MLT2CB repeat: matches 501..455 of consensus"
10208..10679
repeat_region
10208..10679
/note="MIR repeat: matches 494..5 of consensus"
11071..11648
repeat_region
11071..11648
/note="L1 repeat: matches 9..5390 of consensus"
15303..17189
repeat_region
15303..17189
/note="L1P2 repeat: matches 1..891 of consensus"
18419..18722
repeat_region
18419..18722
/note="AluX repeat: matches 1..302 of consensus"
18729..21555
repeat_region
18729..21555
/note="L1 repeat: matches 2533..5384 of consensus"
21419..21991
repeat_region
21419..21991
/note="L1M1 repeat: matches 6..549 of consensus"
21995..22287
repeat_region
21995..22287
/note="AluX repeat: matches 288..1 of consensus"
22288..22817
repeat_region
22288..22817
/note="L1M1 repeat: matches 543..1079 of consensus"
22859..23715
repeat_region
22859..23715
/note="L1M6 repeat: matches 64..922 of consensus"
24116..25064
repeat_region
24116..25064
/note="L1 repeat: matches 4623..3674 of consensus"
25153..25452
repeat_region
25153..25452
/note="AluX repeat: matches 302..1 of consensus"
25455..26328
repeat_region
25455..26328
/note="L1 repeat: matches 3686..2773 of consensus"
27792..27829
repeat_region
27792..27829
/note="L1 repeats of 2 mer 100 % conserved"
28835..29567
repeat_region
28835..29567
/note="MIR repeat: matches 96..250 of consensus"
30233..30545
repeat_region
30233..30545
/note="L1P2 repeat: matches 579..892 of consensus"
30949..31364
repeat_region
30949..31364
/note="L1P3 repeat: matches 896..482 of consensus"
31347..31698
repeat_region
31347..31698
/note="THEIC repeat: matches 2..371 of consensus"
31700..32191
repeat_region
31700..32191
/note="L1M3 repeat: matches 501..1 of consensus"
32043..33837
repeat_region
32043..33837
/note="L1 repeat: matches 5390..3544 of consensus"
33861..34173
repeat_region
33861..34173
/note="Alu repeat: matches 301..1 of consensus"
34500..34538
repeat_region
34500..34538
/note="MIR2 repeat: matches 102..140 of consensus"

```

```

repeat_region 35783..36023
repeat_region /note="L1P4 repeat: matches 648..888 of consensus"
repeat_region 37225..37268
repeat_region /note="22 copies of 2 mer 93 & conserved"
repeat_region 38240..38439
repeat_region /note="MIR repeat: matches 222..32 of consensus"
repeat_region 39158..4010
repeat_region /note="L1M2 repeat: matches 1055..1 of consensus"
repeat_region 40054..40932
repeat_region /note="L1 repeat: matches 5390..4523 of consensus"
repeat_region 40973..41138
repeat_region /note="MIR repeat: matches 191..2 of consensus"
repeat_region 41204..41432
repeat_region /note="MER3 repeat: matches 209..6 of consensus"
repeat_region 41612..41707
repeat_region /note="MIR2 repeat: matches 53..146 of consensus"
repeat_region 43556..43854
repeat_region /note="AluSg repeat: matches 1..300 of consensus"
repeat_region 44857..44908
repeat_region /note="26 copies of 2 mer 92 & conserved"
repeat_region 44909..44985
repeat_region /note="MER34 repeat: matches 456..533 of consensus"
repeat_region 45260..46134
repeat_region /note="L1M7 repeat: matches 917..5 of consensus"
repeat_region 45989..46771
repeat_region /note="L1 repeat: matches 5390..4604 of consensus"
repeat_region 46768..47613
repeat_region /note="L1 repeat: matches 4574..3737 of consensus"
repeat_region 47622..47735
repeat_region /note="MER34 repeat: matches 521..410 of consensus"
repeat_region 47741..48041
repeat_region /note="AluSx repeat: matches 302..1 of consensus"
repeat_region 48042..48495
repeat_region /note="MER34 repeat: matches 431..8 of consensus"
repeat_region 52726..52789
repeat_region /note="L1M2 repeat: matches 1074..1011 of consensus"
repeat_region 52726..52834
repeat_region /note="L1M1 repeat: matches 957..839 of consensus"
repeat_region 52870..53229
repeat_region /note="THE1B repeat: matches 364..1 of consensus"
repeat_region 53233..53532
repeat_region /note="L1M1 repeat: matches 797..507 of consensus"
repeat_region 53785..53906
repeat_region /note="MIR2 repeat: matches 142..23 of consensus"
repeat_region 54257..54327
repeat_region /note="MER33 repeat: matches 1..71 of consensus"
repeat_region 54336..54631
repeat_region /note="L1M1 repeat: matches 66..358 of consensus"
repeat_region 54584..55367
repeat_region /note="L1M3 repeat: matches 256..1051 of consensus"
repeat_region 55374..55539
repeat_region /note="MER33 repeat: matches 53..246 of consensus"
repeat_region 55563..55856
repeat_region /note="AluB repeat: matches 298..1 of consensus"
repeat_region 56799..57352
repeat_region /note="L1P43 repeat: matches 903..353 of consensus"
repeat_region 57364..59701
repeat_region /note="L1 repeat: matches 5297..3008 of consensus"
repeat_region 59701..60002
repeat_region /note="L1M2 repeat: matches 1..304 of consensus"
repeat_region 60033..60406
repeat_region /note="THE1B repeat: matches 364..1 of consensus"
repeat_region 60412..60627
repeat_region /note="L1 repeat: matches 329..553 of consensus"
repeat_region 60470..60627
repeat_region /note="L1M2 repeat: matches 295..453 of consensus"
repeat_region 60624..62683
repeat_region /note="L1 repeat: matches 3017..969 of consensus"
repeat_region 64460..64848
repeat_region /note="MIR2 repeat: matches 1..394 of consensus"
repeat_region 65511..65554
repeat_region /note="FLM1 repeat: matches 1..133 of consensus"
repeat_region 66464..66765

```

```

/note="AluSx repeat: matches 1..302 of consensus"
66789..66830
/note="21 copies of 2 mer 81 & conserved"
67567..67918
/note="L1M1 repeat: matches 19..365 of consensus"
68226..68349
/note="MER34 repeat: matches 189..65 of consensus"
68431..68552
/note="MIR repeat: matches 70..211 of consensus"
68633..68706
/note="L1M2 repeat: matches 1055..983 of consensus"
68697..70419
/note="L1 repeat: matches 3804..2022 of consensus"
70830..71387
/note="L1 repeat: matches 2001..1428 of consensus"
<71389..>71710
/note="match: 3' EST N58388 clone 248070"
72933..73493
/note="L1M2 repeat: matches 553..1 of consensus"
73177..73493
/note="L1M2 repeat: matches 309..1 of consensus"
74118..74451
/note="MER2 repeat: matches 1..345 of consensus"
74959..74996
repeat_region

Query Match 1.38; Score 37.6; DB 10; Length 107677;
Best Local Similarity 52.68; Pred. No. 46;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 2774 tgsagatcactgaccagtagtggtgaagacatatcttctctgtatgtagctgtaga 2833
Db 2524 TGAAGAAAATGGAGACTTTGGCTGCTCTAAATATACATCCCTGTCATCTATTCCAAA 2583

Qy 2834 acagctagtggtgtttttacacagtagtctataaaataggagtggtgataatggc 2893
Db 2584 GGACGCTATTGTTGTTTATCCAAAAGATTTCCAGTAAATGAAGAAATTCAAAAA 2643

Qy 2894 ggtactcatgaacgcgaaatggcgatagatttc 2929
Db 2644 GAGCTTCTAAAACACCATATTCCTGCTGATATAC 2679

RESULT 13
AP000648 151496 bp DNA HTG 04-FEB-2000
LOCUS Homo sapiens chromosome 11 clone CMB9-76D16 map 11q14, WORKING
DEFINITION DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AP000648
VERSION AP000648.1 GI:6997525
KEYWORDS HTG; HNGS_PHASE1; HTGS_DRAFT
SOURCE Homo sapiens DNA, clone: CMB9-76D16.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 151496)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 151,496 genomic DNA of 11q14
Published Only in Database (1999) In press
REFERENCE 2. (bases 1 to 151496)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (28-OCT-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-42-778-9923, Fax: 81-42-778-9924)
On Feb 18, 2000 this sequence version replaced gi:6172413.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN

```


* consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6172: contig of 6172 bp in length
 * gap of unknown length
 * 6173 13947: contig of 7775 bp in length
 * gap of unknown length
 * 13948 21015: contig of 7068 bp in length
 * gap of unknown length
 * 21016 42198: contig of 21183 bp in length
 * gap of unknown length
 * 42199 65368: contig of 23170 bp in length
 * gap of unknown length
 * 65369 93602: contig of 28234 bp in length
 * gap of unknown length
 * 93603 124558: contig of 30956 bp in length
 * gap of unknown length
 * 124559 161247: contig of 36689 bp in length
 * gap of unknown length
 * 161248 200965: contig of 39718 bp in length.

FEATURES Location/Qualifiers
 Source
 1..200965
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-1015N5"
 /chromosome="3p"

BASE COUNT 58185 a 40692 c 41385 g 60699 t 4 others
 ORIGIN

Query Match 1.38; Score 37; DB 53; Length 200965;
 Best Local Similarity 46.18; Pred. NO. 75;
 Matches 124; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
 QY 1427 cacqgaagcttaccggtgtcttattcttcttgcagtgagcgttcagcagggccgg 1486
 DB 20845 CGCCGACGCTTAGAGTCGCGCGCGGCCCACTCCCTCACGGGCCCGCGGCGCAGCGG 20786
 QY 1487 tattaacaggtccattcaccagcagtgatcgagcagcgttcgtcaacttgcacgtc 1546
 DB 20785 CGGCGGAA 20726
 QY 1547 tcaacttaccagtgatgagtgccgcagcagtgccgcagcgttcgttaccacaaa 1606
 DB 20725 TAACCGGGTGATGTTGAGAGGGGTGGCGCTCGGGTGTGCGCGCGCGCGCGCGCGG 20666
 QY 1607 gtggcaggtgcatcttgcacgcccgcagcgtctcgtgggtaccctccacacaccat 1666
 DB 20665 GGGGAAGCAGGACGG 20606
 QY 1667 cgcgggggcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1695
 DB 20605 GGAGGT 20577

Search completed: May 17, 2000, 06:18:16
 Job time: 5954 sec

Query Match					
Best Local Similarity 1.2%; Score 35; DB 1; Length 5280;					
Matches 68; Conservative 0; Mismatches 115; Indels 0; Gaps					
OY	2650	tcaattgttgagagatcacctagtagtggcaggacgtatatctctgtagcgctctgtaagaa	2709		
DB	4218	TCACTTTAGTTGAAGCGGNN	4159		
OY	2710	ettccacaagaccycttacgtyttgttgtttagtcgatcgcctctctgttacttgaccgat	2769		
DB	4158	NNN	4099		
OY	2770	ccattgagatcacctgtaccagatagggtgaagacataattttctccgttatgatctg	2829		
DB	4098	GTGTGGATGTTGGTGCTCCACTTTAGTGGGAAGATAGTGTGTCGCGTTGCTGCTGC	4039		
OY	2830	tag 2832			
DB	4038	TTG 4036			
 RESULT 4					
ID	XO7126	standard; DNA; 799 BP.			
AC	XO7126;				
DE	21-MAY-1999	(first entry)			
DE	Staphylococcus aureus mutant pLIC66	virulence gene.			
KW	virulence; pLIC66;	vaccine; antibacterial; antibiotic;			
KW	screening; bacteremia;	Oligopeptide transporter; ss.			
OS	Staphylococcus aureus				
FH	Key:	Location/Qualifiers			
FT	CD5	1..387			
FT		/tag- a			
PN	WO9901473-A2.				
PP	14-JAN-1999.				
PD	03-JUL-1998;	G01974.			
PPR	03-JUL-1997;	US-887534.			
PR	(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.				
PI	Holiden DW;	99-105999/09.			
P1	WPI;	99-105999/09.			
PT	P-FSDB;	W97716.			
DPT	Inhibition of virulence genes from Staphylococcus aureus				
DPT	for, e.g., screening for potential anti-microbial agents				
PS	Claim 1, page 156-157; 203pp;	English.			
PS	This is the nucleotide sequence of a virulence gene identified in				
CC	Staphylococcus aureus mutant pLIC66. S. aureus genes (see X07088-136).				
CC	associated with virulence were identified by signature-tagged				
CC	magnetesis in which mutants containing a chromosomal insertion of				
CC	a signature tagged transposon were generated, mutants with				
CC	attenuated virulence were identified in a mouse model of bacteraemia,				
CC	and the nucleotide sequences of the regions flanking the transposon				
CC	insertion sites of these mutants were determined. Database sequence				
CC	comparisons were performed to identify the virulence genes and to				
CC	determine the possible function of their protein products (see also				
CC	W97680-724). The pLIC66 virulence gene product (see W97716) is				
CC	unknown, but the gene is located near to the Opp operon. A claimed				
CC	method of identifying an antibacterial agent involves assaying				
CC	potential agents for the ability to interfere with the expression				
CC	of S. aureus virulence gene products. Also new is S. aureus				
CC	containing a functional mutation in one of the virulence genes, and				
CC	its use in vaccine compositions				
SQ	Sequence 799 BP; 283 A; 128 C; 127 G; 259 T;				
 Query Match					
Best Local Similarity 1.2%; Score 34.8; DB 1; Length 799;					
Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps					
OY	340	gggagaatacaactaggttttcacgcagcagcagcgcctatatcaagtacctggaagcct	399		
DB	475	GGCAAAATTACAACTATATTAGACGCAGCTAGATAAATACCATAAGAATTTGGAATGAT	534		
OY	400	ctgagacttggaagaaggtaaagcttccaacgccagcggcatctcgcgtctccaagatcctg	459		

Db	535	TCACCATCAATGAGTATACCAATTGATATACAGCGCAAATAATTAAACAATATACTGC	519
Qy	460	gtaaaattatggccagggtccgcatacgccttggcgtagtggtgactggcccacacct	519
Db	595	ATTACAATATATATTGTATGATCAGATGTTAAATATATATAGCATTTGGTATGACCACAATGA	654
Qy	520	gtttggcattcacagaggtcacagccagtggtgttttccttgagggaacacaagcacac	577
Db	655	TATTGTCATGTTACAACTGCTAGTAGTGCTATATTATATAGGACCACATGAGAAGCATAC	712
<hr/>			
RESULT	5		
ID X0127			
AC	X0127 standard; DNA; 799 BP.		
DT	21-MAY-1999 (first entry)		
DE	Staphylococcus aureus mutant PilC66 virulence gene.		
KW	Virulence; pilC66; vaccine; antibacterial; antibiotic;		
KS	screening; bacteraemia; oligopeptide transporter; ss.		
OS	Staphylococcus aureus.		
FH	Key Location/Qualifiers		
CD5	335..799		
FT	/*tag= a		
FN	WO9901473.A2.		
PD	14-JAN-1999		
PF	03-JUL-1998; G01974		
PR	03-JUL-1997; GS-887534.		
PA	(LMCO) IMPERIAL COLLEGE INNOVATIONS LTD.		
B1	Holten DW; W91717.		
DR	P-FSDS; W91717.		
PT	Inhibition of virulence genes from Staphylococcus aureus - useful		
PS	for, e.g. screening for potential anti-microbial agents		
CC	Claim 1; Page 158; 203pp; English.		
CC	This is the nucleotide sequence of a virulence gene identified in		
CC	Staphylococcus aureus mutant PilC66. S. aureus genes (see X07088-136.)		
CC	associated with virulence were identified by signature-tagged		
CC	mutagenesis in which mutants containing a chromosomal insertion of		
CC	a signature tagged transposon were generated, mutants with		
CC	attenuated virulence were identified in a mouse model of bacteraemia,		
CC	and the nucleotide sequences of the regions flanking the transposon		
CC	insertion sites of these mutants were determined. Database sequence		
CC	comparisons were performed to identify the virulence genes and to		
CC	determine the possible function of their protein products (see also		
CC	W97680-724). The PilC66 virulence gene product (see W9717) is		
CC	unknown, but the gene is located near to the Opp operon. A claimed		
CC	method of identifying an antibacterial agent involves assaying		
CC	potential agents for the ability to interfere with the expression		
CC	of S. aureus virulence gene products. Also new is S. aureus		
CC	containing a functional mutation in one of the virulence genes, and		
CC	its use in vaccine compositions.		
SQ	Sequence 799 BP: 283 A; 128 C; 127 G; 259 T;		
<hr/>			
Query Match			
Best Local Similarity 1.23; Score 34.8; DB 1; Length 799;			
Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps			
Qy	340	gggagaatacataggttttcataccaaggagagcccttatcataagtctgaagcgt	399
Db	475	GCCAANTATCACTATATTGACGAGCTAGTAATATCACATCAAGAATGGAAATGAT	534
Qy	400	ctgagacttgggaagsggttaaagcttccaacgagcgcatactcggtctccaagtacctgg	459
Db	535	TCACCATCAATGAGTATACCAATTGATATACAGCGCAAATAATTAAACAATATACTGC	594
Qy	460	gtaaaattatggccagggtccgcatacgccttggcgtagtggtgactggcccacacct	519
Db	595	ATTACAATATATATTGTATGATCAGATGTTAAATATATATAGCATTTGGTATGACCACAATGA	654
Qy	520	gtttggcattcacagaggtcacagccagtggtgttttccttgagggaacacaagcacac	577

QY 1615 tgcacgattgcacgccccgaggctctcggggggtacctccacaacaccatcgc 1669
 Db 343 TTCCAGGCCCTGGATGCCAGGACTCCCCGcAGGTTGCCCGcCTGCCCCAGCGC 397

```

RESULT      8
V27872
ID      V27872 standard; cDNA; 2848 BP.
AC      AC
DC      12-OCI-1998 (first entry)
DE      Human telomerase protein 2. (TP2) cDNA partial clone 32.
KW      TP2; human; telomerase protein 2; cancer; AIDS; ageing; therapy;
        ss.
OS      Homo sapiens.
Key      Location/Qualifiers
FT      1920..2820
FT      /tag= a
FT      /note= this region includes 7 reverse
FT      transcriptase motifs (Claim 1b)"
FT      misc_feature      1..1689
FT      /tag= b
FT      /note= (Claim 5)"
FT      misc_feature      1..1920
FT      /tag= c
FT      /note= (Claim 5)"
FT      misc_feature      2089..2820
FT      /tag= d
FT      /note= (Claim 5)"
FT      misc_feature      2089..2859
FT      /tag= e
FT      /note= (Claim 5)"

```

WP09221343-31.
P1
P2 12-MAY-1998.
P3 12-NOV-1997. U21348.
P4 16-OCT-1997. US-951733.
P5 15-OCT-1996. US-871189.
P6 11-JUN-1997. US-873039.
P7 (AMGE-) AMGEN CANADA INC.
P8 (AMGE-) AMGEN INC.
P9 HARRISBURG, PA.
P10 HP198-09-20-046.
P11 P1FSDS: W61349.
P12 New nucleic acid encoding human telomerase protein-2 - used for regulating telomerase activity, e.g. for treating cancer or acquired immune deficiency syndrome.
P13 Claim 1, A Fig 5; 15ppr. English.
P14 Partial cDNA clone 32 codes for a large portion (see W61349) of human telomerase protein 2 (Tp2, see W61350), a novel protein of the telomerase complex. It was obtained by PCR amplification (see also V27873-74) of a human colon tumour cell line LIM1863 cDNA. A 3' sequence was subsequently obtained (see V27875), and this was combined with the clone 32 sequence to obtain full-length cDNA (see V27876) for human Tp2. (see W61350). Expressing Tp2 genes in a cell is used to increase telomerase activity and thus proliferation for treatment of e.g. HIV infection, AIDS and ageing disorders, while expressing an inactive mutant of Tp2 (or molecule antisense to the gene) is used to decrease telomerase activity, e.g. for treatment of cancer. Tp2 polypeptides can also be used to screen for agents that inhibit Tp2 activity or its binding to TRIP1 (see W61347) or telomerase RNA, potentially useful therapeutically, also to raise specific antibodies useful in immunoassays and therapeutically as inhibitors. Nucleic acid fragments are used as diagnostic probes for detecting/quantifying Tp2 DNA. Also contemplated are transgenic animals in which the Tp2 gene has been inactivated or is overexpressed. Tp2 polypeptides are administered i.v., s.c. or orally, or they are delivered from engineered cells or gene therapy vectors.

Sequence 2848 BP: 437 A; 945 G; 488 T;

Query Match 1.2%; Score 34.2; DB 1; Length 2848;
Best Local Similarity 49.7%; Pred. No. 2;

Matches	87;	Conservative	0;	Mismatches	88;	Indels	0;	Gaps	0;
QY	1495	gtccattcgcacagtatcgccaggcgggctctgccaatttcggcacgtctcaact	1554						
Db	1047	GACCAAGCATCTCCTACTCTCCTCAGCGGACAAGAGAGTGGCGGCCCTTCCTACT	1106						
QY	1555	teagattgatagtgccgccagagtcggcgagcgtcctcgtttaccaaaadggcagg	1614						
Db	1107	CAGCTCTCTCAGGCCCCAGCCCTGACTGCCTCGGAGGTCTGGAGACCAATCTTTCTGG	1166						
QY	1615	tgcatacgttgcagcccccgaggtcctcgggggtacctccacaacacattcgc	1669						
Db	1167	TTCAGGCCCTTGATGCCAGGAGACTCCCGCAGGTTGCCCGCCCTGCCCCAGCGC	1221						
RESULT	9								
V27876									
ID	V27876	standard; cDNA; 3798 BP.							
AC	V27876;								
CD	12-OCT-1998	(first entry)							
DE	Human telomerase protein 2 (TP2) full-length cDNA.								
DT	TP2; human; telomerase protein 2; cancer; AIDS; ageing; therapy;								
KW	ss.								
OS	Homo sapiens.								
HF	Key	Location/Qualifiers							
FH	2 . 3466								
FT	/ *tag= a								
PD	WO9821343-A1.								
PN	22-MAY-1998.								
PF	13-NOV-1997; U21248.								
PR	16-OCT-1997; US-951733.								
PR	15-NOV-1996; US-871189.								
PR	11-JUN-1997; US-873039.								
PA	(AMGE-) AMGEN CANADA INC.								
PI	(AMGE-) AMGEN CAN.								
PI	Harrington LA; Robinson MO;								
PT	WPI; 98-297946/26.								
PT	p-P5DB; n27876								
PP	New nucleic acid encoding human telomerase protein-2 - used for								
PT	regulating telomerase activity, e.g. for treating cancer or acquired								
PT	immune deficiency syndrome								
PS	Claim 1c; Fig 8; 150pp; English.								
CC	This full-length cDNA encodes for a human telomerase protein								
CC	2 (TP2; see W61350), a novel protein of the telomerase complex								
CC	The sequence was deduced from overlapping partitions #32 (see								
CC	V27876) and TP2-15 (see V27875) which were obtained from a human								
CC	colon tumour cell line LM1863 cDNA. Expressing TP2 genes in a								
CC	cell is used to increase telomerase activity and thus proliferation								
CC	for treatment of e.g. HIV infection, AIDS and ageing disorders,								
CC	while expressing an inactive mutant of TP2 (or molecule antisense								
CC	to the gene) is used to decrease telomerase activity, e.g. for								
CC	treatment of cancer. TP2 polypeptides can also be used to screen								
CC	for agents that inhibit TP2 activity or its binding to TRIP1 (see								
CC	W61347) or telomerase RNA, potentially useful therapeutically,								
CC	also to raise specific antibodies useful in immunoassays and								
CC	therapeutically as inhibitors. Nucleic acid fragments are used								
CC	as diagnostic probes for detecting/quantifying TP2 DNA. Also,								
CC	contemplated are transgenic animals in which the TP2 gene has been								
CC	inactivated or is overexpressed. TP2 polypeptides are administered								
CC	i.v., s.c. or orally, or they are delivered from engineered cells								
CC	or gene therapy vectors.								
SQ	Sequence 3798 BP:	613 A;	1311 C;	1212 G;	662 T;				

```

Query Match      1.28; Score 34.2; DB 1; Length 3798;
Best Local Similarity 49.7%; Pred. No. 2.4;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1495 gtccattaccgacagtatcggcaggcggcgtcgtcaattcggcagctctccact 1554
      |||||      ||      ||      ||      ||      ||      ||      ||
Db 1048 GACCAAGCATCTCTCTACTCTCTCTCAAGGAGCAGCTGGGCCCTCTCTACT 1107

QY 1555 teagatgatagggtgcgcgcagctggcgcagcgtcgtctctgtaccaaaatggcagg 1614

```

```

Query Match      1.28; Score 34.2; DB 1; Length 3798;
Best Local Similarity 49.7%; Pred. No. 2.4;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1495 gtccattccgcagagtatcggcaggcgggctgtgtaattcggcagctctcctaact 1554
      |||||      ||      ||      ||      ||      ||      ||      ||
Db 1048 GACCAAGCATCTTCTCTACTCTCTCTCAAGGAGGAGCTGTGGGCCCTCTTCTACT 1107

QY 1555 teagatgatagggtgcgcgcagctggcgcagcgtgctctgtttaccaaaatggcagg 1614

```

Db 1108 CAGCTCTGAGGCGCCAGCTGCGCTCGAGGCTCGGAGGACCTTCCTGGG 1167
 Qy 1615 tgcacgattgcacgcccagagctctcggggtaccctccacacaccatgc 1669
 Db 1168 TTCAGGCGCTGATGCGAGGACTCCCGCAGGTGCGCGCTGCCAGGCGC 1222
 RESULT 10
 V22382
 ID V22382 standard; cDNA: 3855 BP.
 AC V22382;
 DT 13-AUG-1998 (first entry)
 DE Human telomerase reverse transcriptase Delta182 variant encoding cDNA.
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
 OS Synthetic.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 56..2479
 FT /tag= a
 FT /product= "telomerase reverse transcriptase variant"
 FT
 PN GB2317891-A.
 PD 08-APR-1998.
 PF 01-OCT-1997; 020890.
 PR 14-AUG-1997; US-915503.
 PR 01-OCT-1996; US-724643.
 PR 18-APR-1997; US-844419.
 PR 25-APR-1997; US-846017.
 PR 06-MAY-1997; US-851843.
 PR 09-MAY-1997; US-854050.
 PR 14-AUG-1997; US-911312.
 PR 14-AUG-1997; US-912951.
 PA (GERO-) GERON CORP.
 PA (UTTE-) UNIV TECHNOLOGY CORP.
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,
 PI Morin GB, Nakamura T, Harley CB;
 DR WPI: 98-171633/16.
 DR P-PSDB: W46997.
 DR Pure and recombinant human Telomerase Reverse Transcriptase and its
 PT variants - are useful in the diagnosis, prognosis and treatment of
 cell proliferation conditions especially cancer and aging
 PS Disclosure: Fig 20; 387pp; English.
 CC The present sequence encodes a human telomerase reverse transcriptase
 CC (hTERT) variant from the present invention. The present invention also
 CC describes the following methods: (A) determining whether a test compound
 CC is a modulator of hTERT by detecting the change in hTERT recombinant
 CC preparation of recombinant telomerase by contacting a protein
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the
 CC sample and detecting the complex formed or in the case of RNA detection,
 CC amplifying the product with presence of hTERT in the sample; and (D)
 CC increasing the proliferation of a vertebrate cell by increasing hTERT
 CC expression; and (E) the use of an agent that causes an increase in cell
 CC vertebrate cell proliferation to create a medicament that inhibits
 CC aging. A protein preparation of hTERT and the polynucleotide encoding
 CC hTERT can be used in the manufacture of medicaments for inhibiting the
 CC effect of aging or cancer. Inhibitors of telomerase activity can be
 CC used to treat conditions that are associated with high telomerase
 CC activity. A protein preparation of hTERT can also be used in the new
 CC methods.
 SQ Sequence 3855 BP; 651 A; 1300 C; 1226 G; 678 T;

Query Match 1.2%; Score 34.2; DB 1; Length 3855;
 Best Local Similarity 49.7%; Pred. No. 2.4;
 Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 Qy 1495 gtccattcatcagcagtgatcgagggcggttcgtcaacttcgacgtctactact 1554
 Db 1036 GACCAAGACATTCCTCTACTCTCCTCAGGCGACAGGAGGAGGCGGCGCTCTCTACT 1095

Qy 1555 tcagattgatagtgccgcccagctgagcgagtgctcgtgttaccacaaagtgagcag 1614
 Db 1096 CAGCTCTCTGAGGCGCCAGCTGCGCTCGAGGCTCGGAGGACCTTCCTGGG 1155
 Qy 1615 tgcacgattgcacgcccagagctctcggggtaccctccacacaccatgc 1669
 Db 1156 TTCAGGCGCTGATGCGAGGACTCCCGCAGGTGCGCGCTGCCAGGCGC 1210
 RESULT 11
 V22379
 ID V22379 standard; cDNA: 3955 BP.
 AC V22379;
 DT 13-AUG-1998 (first entry)
 DE Human telomerase reverse transcriptase encoding cDNA.
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
 OS Synthetic.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 56..3454
 FT /tag= a
 FT /product= "telomerase reverse transcriptase"
 FT
 PN GB2317891-A.
 PD 08-APR-1998.
 PF 01-OCT-1997; 020890.
 PR 14-AUG-1997; US-915503.
 PR 01-OCT-1996; US-724643.
 PR 18-APR-1997; US-844419.
 PR 25-APR-1997; US-846017.
 PR 06-MAY-1997; US-851843.
 PR 09-MAY-1997; US-854050.
 PR 14-AUG-1997; US-911312.
 PR 14-AUG-1997; US-912951.
 PA (GERO-) GERON CORP.
 PA (UTTE-) UNIV TECHNOLOGY CORP.
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,
 PI Morin GB, Nakamura T, Harley CB;
 DR WPI: 98-171633/16.
 DR P-PSDB: W46997.
 DR Pure and recombinant human Telomerase Reverse Transcriptase and its
 PT variants - are useful in the diagnosis, prognosis and treatment of
 cell proliferation conditions especially cancer and aging
 PS Claim 5; Fig 16; 387pp; English.
 CC The present sequence encodes human telomerase reverse transcriptase
 CC (hTERT), which is a ribonucleoprotein. The present invention also
 CC describes the following methods: (A) determining whether a test compound
 CC is a modulator of hTERT by detecting the change in hTERT recombinant
 CC preparation of recombinant telomerase by contacting a protein
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the
 CC sample and detecting the complex formed or in the case of RNA detection,
 CC amplifying the product with presence of hTERT in the sample; and (D)
 CC increasing the proliferation of a vertebrate cell by increasing hTERT
 CC expression; and (E) the use of an agent that causes an increase in cell
 CC vertebrate cell proliferation to create a medicament that inhibits
 CC aging. A protein preparation of hTERT and the polynucleotide encoding
 CC hTERT can be used in the manufacture of medicaments for inhibiting the
 CC effect of aging or cancer. Inhibitors of telomerase activity can be
 CC used to treat conditions that are associated with high telomerase
 CC activity. A protein preparation of hTERT can also be used in the new
 CC methods.
 SQ Sequence 3955 BP; 648 A; 1353 C; 1251 G; 703 T;

Query Match 1.2%; Score 34.2; DB 1; Length 3955;
 Best Local Similarity 49.7%; Pred. No. 2.5;
 Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 Qy 1495 gtccattcatcagcagtgatcgagggcggttcgtcaacttcgacgtctactact 1554
 Db 1036 GACCAAGACATTCCTCTACTCTCCTCAGGCGACAGGAGGAGGCGGCGCTCTCTACT 1095

```

Db 1036 GACCAAGCACTTCTCTACTCTCTCAGGCGACAAGGAGCAGCTGCGGCCCTCTCTCTACT 1095
QY 1555 tcagattgatagtcgcccagcgtgcccagcgtgctcgtgtttaccacaaagtggcagg 1614
Db 1096 CAGCTCTCTGAGGCCAGCCTGACTTGGCGCTCGGAGGCTGTGGAGACCACTTTCCTGGG 1155
QY 1615 tgcatcgatttcacgcccagcgtctcgggggtaccctcccaaacaccatcgc 1669
Db 1156 TTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTCCCGCCCTGCCCGCCGCGC 1210

RESULT 12
V72125
ID V72125 standard; cDNA; 3972 BP.
AC V72125
DE Human catalytic telomerase sub-unit RACE-generated cDNA #4.
KW Human; catalytic telomerase subunit; therapy; diagnosis; htc; assay;
KW modulator; treatment; inhibit; cellular disorder; death; defect; cancer;
KW ageing; antisense; neoplastic cell; telomerase-related condition; RACE;
KW tumour cell; ss.
OS Homo sapiens.
FN WO9859040-A2.
PD 30-DEC-1998.
PF 09-JUN-1998; E03468.
PR 14-APR-1998; DE-016496.
PR 20-JUN-1997; DE-026329.
PR 26-MAR-1998; DE-013274.
PA (FARB ) BAYER AG
PI Hagen G, Siegmund H, Weichel W, Wick M, Zubov D;
DR WPI; 99-0812/6/07.
DE New catalytically active subunit of human telomerase - used in the
PT modulation of telomerase activity, particularly for treating cancer
PT and ageing
PS Example 6; Fig 11; 76pp; German.
CC This sequence represents a RACE-generated fragment of a novel human
CC catalytic telomerase sub-unit (hTC). The hTC protein can be used
CC in screening assays to identify modulators of telomerase and to treat or
CC inhibit cellular disorders, death, defects and/or other pathological
CC processes involving telomerase, particularly cancer and ageing (also
CC suitable for this are agents that stimulate, inhibit or mimic the
CC activity of the subunit). Antisense nucleic acids inhibit telomerase
CC action (by binding to specific mRNA), particularly in neoplastic cells
CC and may be expressed in vivo. Antibodies and fragments of the protein,
CC used as probes or primers, are used to diagnose telomerase-related
CC conditions (especially neoplasia) by (i) detecting abnormal levels of
CC the subunit protein in body fluids or tissues or (ii) by measuring the
CC amount of the encoding nucleic acid. Expression of the nucleic acid
CC encoding the subunit mRNA is confined to tumour cells, in contrast to
CC the ubiquitous expression of the telomerase RNA subunit.
CC Sequence 3972 BP; 701 A; 1287 C; 1264 G; 720 T;

Query Match 1.2%; Score 34.2; DB 1; Length 3972;
Best Local Similarity 49.7%; Pred. No. 2.5;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1495 gtccattcatcagcagtgatcgccagcgtgcccagcgttctcgaacttcgacgtctcactact 1554
Db 1063 GACCAAGCACTTCTCTACTCTCTCAGGCGACAAGGAGCAGCTGCGGCCCTCTCTCTACT 1122
QY 1555 tcagattgatagtcgcccagcgtgcccagcgtgctcgtgtttaccacaaagtggcagg 1614
Db 1123 CAGCTCTCTGAGGCCAGCCTGACTTGGCGCTCGGAGGCTGTGGAGACCACTTTCCTGGG 1182
QY 1615 tgcatcgatttcacgcccagcgtctcgggggtaccctcccaaacaccatcgc 1669
Db 1183 TTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTCCCGCCCTGCCCGCCGCGC 1237

RESULT 13
V60320
ID V60320 standard; cDNA; 4023 BP.

```

```

AC V60320:
DE 04-DEC-1998 (first entry)
DE Human telomerase gene referred to as hEST2.
DE Catalytic subunit; human telomerase; telomere maintenance;
DE diagnosis; treatment; cancer; ss.
KW Homo sapiens.
KW Key Location/Qualifiers
FT CDS 59..3458
FT FT /*tag= a
PN WO9837181-A2.
PD 27-AUG-1998.
PF 20-FEB-1998; J03404.
PF 30-OCT-1997; US-064322.
PR 20-FEB-1997; US-038750.
PR 20-MAY-1997; US-047151.
PR 01-AUG-1997; US-054549.
PR 14-AUG-1997; US-055762.
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA Counter CM, Meyerson M, Weinberg RA;
PI WPI; 98-495367/42.
DR P-PSDB; W71376.
DE New isolated human telomerase catalytic sub-unit gene - used to
DE develop products for increasing or reducing the life span of cells
DE such as cancer cells or transformed cells
PS Claim 5; Fig 5A-B; 96pp; English.
CC The present sequence encodes the catalytic subunit of a human
CC telomerase holoenzyme. Disruption of the telomerase gene alters
CC telomere maintenance. The DNA is essential for telomerase activity,
CC and the protein is physically associated with telomerase and a
CC constituent of active telomerase complex. The products can be used
CC for increasing or reducing the lifespan of cells such as cancer cells
CC or transformed cells. They can also be used in the diagnosis and
CC treatment of malignancies. In addition, cells with a longer lifespan
CC can be transplanted into or grafted onto an individual (e.g. as skin
CC grafts, as systems for delivery of therapeutic proteins, such as hormones
CC and enzymes), to whom they provide therapeutic benefit.
CC Sequence 4023 BP; 668 A; 1363 C; 1277 G; 715 T;

Query Match 1.2%; Score 34.2; DB 1; Length 4023;
Best Local Similarity 49.7%; Pred. No. 2.5;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1495 gtccattcatcagcagtgatcgccagcgtgcccagcgttctcgaacttcgacgtctcactact 1554
Db 1039 GACCAAGCACTTCTCTACTCTCTCAGGCGACAAGGAGCAGCTGCGGCCCTCTCTCTACT 1098
QY 1555 tcagattgatagtcgcccagcgtgcccagcgtgctcgtgtttaccacaaagtggcagg 1614
Db 1099 CAGCTCTCTGAGGCCAGCCTGACTTGGCGCTCGGAGGCTGTGGAGACCACTTTCCTGGG 1158
QY 1615 tgcatcgatttcacgcccagcgtctcgggggtaccctcccaaacaccatcgc 1669
Db 1159 TTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGCCCTGCCCGCCGCGC 1213

RESULT 14
V22428
ID V22428 standard; cDNA; 4037 BP.
AC V22428;
DE 13-AUG-1998 (first entry)
DE Human telomerase reverse transcriptase encoding cDNA refined sequence.
DE Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.
KW Key Location/Qualifiers
FT CDS 56..3454
FT FT /*tag= a
FT FT /product= "telomerase reverse transcriptase"
FT FT /note= "refined sequence"
PN GB2317891-A.
PD 08-APR-1998.
PF 01-OCT-1997; 020890.

```

Thu May 18 22:21:27 2000

us-09-151-189-1.rni

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:39:38 ; Search time 56.46 Seconds
(without alignments)
6591.506 Million cell updates/sec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcgcacagccttggaga.....aaatggcgatagattccc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	1.2	7218	1	US-08-232-463-14
2	34.2	1.2	50341	2	US-08-247-901C-1
3	34.2	1.2	50341	4	US-09-075-904-1
C 4	32	1.1	704	2	US-08-463-115-29
C 5	32	1.1	704	2	US-08-463-388-29
C 6	31.2	1.1	1998	7	US212296-8
7	31	1.1	1150	1	US-08-264-861A-10
8	31	1.1	1150	6	PCT-US95-07784-10
9	31	1.1	2156	4	US-08-959-011-2
10	31	1.1	49272	2	US-08-614-770A-1
11	30.8	1.1	813	1	US-08-276-919-9
12	30.8	1.1	813	2	US-08-776-088-12
13	30.8	1.1	813	6	PCT-US95-09145A-12
14	30.8	1.1	822	2	US-08-776-088-14
15	30.8	1.1	822	6	PCT-US95-09145A-14
16	30.8	1.1	986	1	US-08-276-919-3
17	30.8	1.1	986	1	US-08-276-919-12
18	30.8	1.1	986	2	US-08-776-088-3
19	30.8	1.1	986	2	US-08-776-088-17
20	30.8	1.1	986	6	PCT-US95-09145A-3
21	30.8	1.1	1104	1	PCT-US95-09145A-17
22	30.8	1.1	1104	1	US-08-276-919-1
23	30.8	1.1	1104	2	US-08-776-088-1
24	30.8	1.1	1104	6	PCT-US95-09145A-1
25	30.8	1.1	1438	4	US-08-845-742-1
26	30.8	1.1	2134	2	US-08-776-088-5
27	30.8	1.1	2134	6	PCT-US95-09145A-5

28	30.3	1.1	9515	2	US-08-920-812-13	Sequence 13, Appl
29	30.8	1.1	9515	2	US-08-920-827-13	Sequence 13, Appl
30	30.8	1.1	9515	2	US-08-921-177-13	Sequence 13, Appl
31	30.8	1.1	9515	2	US-08-921-177-13	Sequence 13, Appl
32	30.8	1.1	9515	2	US-08-921-177-13	Sequence 13, Appl
33	30.2	1.0	336	4	US-08-920-828-13	Sequence 3, Appl
34	30.2	1.0	336	4	US-07-814-220-3	Sequence 3, Appl
35	30.2	1.0	336	4	US-07-814-220-4	Sequence 3, Appl
36	30.2	1.0	336	4	US-07-812-421-3	Sequence 3, Appl
37	30.2	1.0	336	4	US-07-812-421-4	Sequence 3, Appl
38	30.2	1.0	1290	5	US-08-745-977-3	Sequence 3, Appl
39	30	1.0	1290	5	US-09-040-699A-3	Sequence 3, Appl
40	30	1.0	434	5	US-08-866-340-25	Sequence 25, Appl
41	30	1.0	1635	1	US-08-133-347-1	Sequence 1, Appl
42	30	1.0	1635	1	US-08-133-347-3	Sequence 3, Appl
43	29.6	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
44	29.6	1.0	1662	2	US-08-651-574-1	Sequence 1, Appl
45	29.6	1.0	1662	5	US-09-066-544-1	Sequence 1, Appl
			2277	2	US-08-676-967-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1.
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9pt-F15
; US-08-232-463-14

[illegible]

RESULT 2
US-08-247-901C-1
Sequence 1, Application US/08247901C
Patent No. 5750384
GENERAL INFORMATION:
APPLICANT: Jacobs et al
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Eberstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247.901C
FILING DATE: May 23, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/057,531
FILING DATE: April 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Bogoslian, Elizabeth A
REGISTRATION NUMBER: 39,911
REFERENCE/POCKET NUMBER: 96700/273
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 50341
TYPE: nucleic acid
STRANDEDNESS: single

```

TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: L5 shuttle phasmid sequence
HYPOTHETICAL: No
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: L5 mycobacteriophage
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5750384e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1

Query Match          1.28; Score 34.2; DB 2; Length 50341;
Best Local Similarity 50.9%; Pred. No. 8.8;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1296 gtcgatttgagaaatgggtgaacattcagaatggtgaagcccccgcgcgcccgcaagag 1355
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db 8714 CTCGACTACGTCCCGAGCGCAGGCCGCTGGAGGGTTTCGCCTCGCGCATGCCGACGAG 8773
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
QY 1356 tttagcaagaagctgcgttttatccgtaaacgcgcacctggccagggttctcttcgtg 1415
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db 8774 GCTGACCAAGAGCTGTGGCAGTGGTGTCAGGCCCAACAACCTCGACATCGAGGACCACCTG 8833
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
QY 1416 gcaataatcaaccgaagcttatcgcggtctcttate 1454
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db 8834 GCCTACACCGACGCTACGTTTCACGCGCGGTCTGCTTCATC 8872
      |||||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||

RESULT 3
US-09-075-904-1
Sequence 1, Application US/09075904
Patent No. 5984137
GENERAL INFORMATION:
APPLICANT: Jacobs, et al.
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible

```

OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Processor (ASCII)
 CURRENT APPLICATION DATA: US/09/075,904
 FILING DATE: May 11, 1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/247,901
 FILING DATE: May 23, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Bogosian, Elizabeth A
 REGISTRATION NUMBER: 39,911
 REFERENCE/DOCKET NUMBER: 96700/475
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-5995
 TELEFAX: (212) 286-0854 or 286-0082
 TELEX: TWX 710-581-4766
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 50341
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: L5 shuttle plasmid sequence
 HYPOTHEICAL: No
 ANTI-SENSE:
 FRAGMENT TYPE:
 ORGANISM: L5 mycobacteriophage
 STRAIN:
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE:
 HAPLOTYPE:
 TISSUE TYPE:
 CELL TYPE:
 CELL LINE:
 ORGANELLE:
 IMMEDIATE SOURCE:
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT:
 FEATURE:
 NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION:
 PUBLICATION INFORMATION: No. 5994137e
 AUTHORS:
 TITLE:
 JOURNAL:
 VOLUME:
 PAGES:
 DATE:
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO:

US-09-075-904-1

Query Match 1.28; Score 34.2; DB 4; Length 50341;
 Best Local Similarity 50.98; Pred. No. 8.8;
 Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 Qy 1296 gtccgatttgacgaatgctgaacattcagaatggtggaccctcccgccgcccgaagag 1355
 Db 8714 GTGACTCTAGTCGCGAGCGCCAGCGCGTTCGAGGCTTCGCGCTCGCGGATGCGCGAG 8773
 Qy 1356 ttgacgaagagctgctgtttatctcgttaacgcccgcgacctggccaggggtctcctctg 1415
 Db 8774 GCTGACGAGAGCTGTGGCAGTGTGGCAGGCGCAACACCTCGACATCGAGCGACCACTG 8833
 Qy 1416 gacaatacaacaccgaagcttatcgcggtctcttctatc 1454

Db 8834 GGCTACCGGAGCTTACGTCACGGCGGCTGCTTCATC 8872
 RESULT 4
 US-08-463-115-29/c
 Sequence 29, Application US/08463115
 Patent No. 5703221
 GENERAL INFORMATION:
 APPLICANT: WILLIAM JOHN MARTIN
 TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
 TITLE OF INVENTION: AND RELATED VACCINES
 NUMBER OF SEQUENCES: 104
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LYON & LYON
 STREET: 633 West Fifth Street
 CITY: Suite 4700
 STATE: Los Angeles
 COUNTRY: California
 ZIP: 90071-2066
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: FastSeq Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,115
 FILING DATE: June 5, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION DATA: including application
 PRIOR APPLICATION DATA: described below:
 APPLICATION NUMBER: 08/157,811
 FILING DATE: No. 5703221ember 23, 1993
 APPLICATION NUMBER: 07/887,502
 FILING DATE: May 22, 1992
 APPLICATION NUMBER: 07/704,814
 FILING DATE: May 23, 1991
 APPLICATION NUMBER: 07/763,039
 FILING DATE: September 20, 1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Warburg, Richard J.
 REGISTRATION NUMBER: 32,927
 REFERENCE/DOCKET NUMBER: 213/301
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 704 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 OTHER INFORMATION:
 US-08-463-115-29

Query Match 1.18; Score 32; DB 2; Length 704;
 Best Local Similarity 57.7%; Pred. No. 2.2;
 Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
 Qy 1757 tcccaacaagaggtcaccttaoctttctccacaagctatccaagtggatcgccaacga 1816
 Db 215 TCCCCCCACGGTNTGACGGCTGCTCCCATAGGTTGCGGTTGCCCATCCCC 156
 Qy 1817 cctctctaccggtccggccagcgctaccacaaatgga 1853
 Db 155 CGTTTCTACCCACGGCGCCGCTCTCCGATTGCA 119

```

RESULT 5
US-08-465-388-29/c
; Sequence 29, Application US/08465388
; Patent No. 5753488
; GENERAL INFORMATION:
; APPLICANT: WILLIAM JOHN MARTIN
; TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,388
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/157,811
; FILING DATE: No. 5753488, 23, 1993
; APPLICATION NUMBER: 07/887,502
; FILING DATE: May 22, 1992
; APPLICATION NUMBER: 07/704,814
; FILING DATE: May 23, 1991
; APPLICATION NUMBER: 07/763,039
; FILING DATE: September 20, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Harburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 213/300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION:
;
US-08-465-388-29

Query Match 1.18; Score 32; DB 2; Length 704;
Best Local Similarity 57.78; Pred. No. 2.2; Indels 0;
Matches 56; Conservative 0; Mismatches 41; Gaps 0;

QY 1757 tcccaacacgaggtccacctcttcccaacagctatcccaagtggatcgcccaacgca 1816
Db 215 TCCCCCGCCGGTCTCCAGCGCTCTCCCATAGGGTGTGCTCGCGGTTCCCGATCCCG 156
QY 1817 ccttctaccctccggtccgacgctaccccaaatgga 1853
Db 155 CGTTTCTACCCAGCGGCCCGCGCTCTCCGATTGCA 119

RESULT 6
5212296-8/c

Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO: 8
; LENGTH: 1998
; 5212296-8

Query Match 1.18; Score 31.2; DB 7; Length 1998;
Best Local Similarity 47.18; Pred. No. 8.5; Indels 0;
Matches 96; Conservative 0; Mismatches 108; Gaps 0;

QY 516 cctgttttgcatcgcagaggtcacagccagtggttttccctgaggaaacacacac 575
Db 1930 CGCTTCTGGCGCGCGCAGCATCAAGGACAAAGGTTTCGGGACGAGCAGCTGGACCAC 1871
QY 576 accggtggaggagagacacctccagacctgtacaaactccgacgacgctggtccgacg 635
Db 1870 CCGGTGTGGCGCGCTGACCTGACCTACGGGACGCTGCGGTGCGCGGCGGACCGGG 1811
QY 636 gcgcgcgaatagaaggagacacgttagcttgcgcgcgcgcgcgcgcgcgcgcgcgc 695
Db 1810 CAGAGCGTGTGACCCACCCAGCGCGTCCGCGGTTCGTCGCGAGGCGCGCTCGGGCA 1751
QY 696 cgtgacgagcagggagctgtctgc 719
Db 1750 CTGGCGCGCGCGCGACGCGGTGC 1727

RESULT 7
US-08-264-861A-10
; Sequence 10, Application US/08264861A
; Patent No. 5622866
; GENERAL INFORMATION:
; APPLICANT: MOTAMED, HAIDER
; APPLICANT: SHAFIEE, ALI
; TITLE OF INVENTION: EXPRESSION CASSETTES USEFUL IN
; TITLE OF INVENTION: CONSTRUCTION OF INTEGRATIVE AND REPLICATIVE EXPRESSION
; TITLE OF INVENTION: VECTORS FOR STREPTOMYCES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. ERIC THIES
; STREET: P.O. BOX 2000, 126 E. LINCOLN AVENUE
; CITY: RAYWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,861A
; FILING DATE: 23-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: THIES, J. ERIC
; REGISTRATION NUMBER: 35,382
; REFERENCE/DOCKET NUMBER: 19132
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (908)594-3904
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-264-861A-10

Query Match 1.1%; Score 31; DB 1; Length 1150;
Best Local Similarity 52.8%; Pred. No. 6.7;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1153 ctttccagctctccgagcagccttctgtgttggttgaaacggggccctttgtctctc 1212
DB 1007 CGATCTCCCTGTTCTGACGCGCCGAGCGCGCCAGCGCGCCAGCGCGCTCCGACTCCG 1066
QY 1213 agctcgtgtgaacagttaccatcagcagctattacggtcgaaacagcagagacat 1272
DB 1067 TCAGTGACGCGATCGCTGCTCCAGCGTCCGTCAGCGCCCGCGATGTCGAGGGCT 1126
QY 1273 tgcgcc 1279
DB 1127 CCGCGCC 1133

RESULT 8
PCT-US95-07784-10
Sequence 10, Application PC/TUS9507784
GENERAL INFORMATION:
APPLICANT: MOTAMED, HAIDEH
APPLICANT: SHAFIEE, ALI
TITLE OF INVENTION: EXPRESSION CASSETTES USEFUL IN
TITLE OF INVENTION: CONSTRUCTION OF INTEGRATIVE AND REPLICATIVE EXPRESSION
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: J. ERIC THIES
STREET: P. O. BOX 2000, 136 E. LINCOLN AVENUE
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07784
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: THIES, J. ERIC
REGISTRATION NUMBER: 35,382
REFERENCE/DOCKET NUMBER: 19132
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)594-3904
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

PCT-US95-07784-10

Query Match 1.1%; Score 31; DB 6; Length 1150;
Best Local Similarity 52.8%; Pred. No. 6.7;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1153 ctttccagctctccgagcagccttctgtgttggttgaaacggggccctttgtctctc 1212
DB 1007 CGATCTCCCTGTTCTGACGCGCCGAGCGCGCCAGCGCGCTCCGACTCCG 1066
QY 1213 agctcgtgtgaacagttaccatcagcagctattacggtcgaaacagcagagacat 1272
DB 1067 TCAGTGACGCGATCGCTGCTCCAGCGTCCGTCAGCGCCCGCGATGTCGAGGGCT 1126
QY 1273 tgcgcc 1279
DB 1127 CCGCGCC 1133

RESULT 9
US-08-959-011-2/c
Sequence 2, Application US/08959011
Patent No. 5932444
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VACUOLAR ATPASE SUBUNIT AC45
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/959,011
FILING DATE: Herewith
CLASSIFICATION: 336
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0412 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: CRBLNOT01
CLONE: 676592
US-08-959-011-2

Query Match 1.1%; Score 31; DB 4; Length 2156;
Best Local Similarity 46.9%; Pred. No. 10;
Matches 97; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Sequence 12, Application US/08776088
Patent No. 5773579
GENERAL INFORMATION:
APPLICANT: Torczynski, Richard M.
APPLICANT: Bollon, Arthur P.
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIDLEY & AUSTIN
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 19-JUL-95
APPLICATION NUMBER: US/08/776,088
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Eugenia S. Hansen
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 10365/05011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 813 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..813
US-08-776-088-12

Query Match 1.1%; Score 30.8; DB 2; Length 813;
Best Local Similarity 49.4%; Pred. No. 6.1;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1331 tggaccgccgccccgagaggttagacgaagagctgctttatccgttaacgccg 1390
DB 306 TGAGCACACCGTCAGCGGACAGCACTTCGCGCGAGCTGCACATTGTCCATTATAACTC 365
QY 1391 cgacctggccagggtctctctcttggaacaataataacacgcagccttatcgcggtctct 1450
DB 366 AGACCTTTATCTCTGAGGAGCTTCGCGGAGCTGCACGAAGTCAGAGGCTGCTCTCTGC 425
QY 1451 tatctacttgagctggagccttcagcagcccggtatcaa 1492
DB 426 TGTTCTCATTGAGATGGGCTCTCTCAATCCGCTCTATGACAA 467

RESULT 13
US-08-776-088-12
Sequence 12, Application PC/TUS9509145A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US

ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 19-JUL-95
APPLICATION NUMBER: PC/TUS95/09145A
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Eugenia S. Hansen
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: 83579ACIPPT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 813 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..813
PCT-US95-09145A-12

Query Match 1.1%; Score 30.8; DB 6; Length 813;
Best Local Similarity 49.4%; Pred. No. 6.1;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1331 tggaccgccgccccgagaggttagacgaagagctgctttatccgttaacgccg 1390
DB 306 TGAGCACACCGTCAGCGGACAGCACTTCGCGCGAGCTGCACATTGTCCATTATAACTC 365
QY 1391 cgacctggccagggtctctctcttggaacaataataacacgcagccttatcgcggtctct 1450
DB 366 AGACCTTTATCTCTGAGGAGCTTCGCGGAGCTGCACGAAGTCAGAGGCTGCTCTCTGC 425
QY 1451 tatctacttgagctggagccttcagcagcccggtatcaa 1492
DB 426 TGTTCTCATTGAGATGGGCTCTCTCAATCCGCTCTATGACAA 467

RESULT 14
US-08-776-088-14
Sequence 14, Application US/08776088
Patent No. 5773579
GENERAL INFORMATION:
APPLICANT: Torczynski, Richard M.
APPLICANT: Bollon, Arthur P.
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIDLEY & AUSTIN
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 19-JUL-95
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Eugenia S. Hansen
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 10365/05011
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..822
US-08-776-088-14

Query Match 1.1%; Score 30.8; DB 2; Length 822;
Best Local Similarity 49.4%; Pred. No. 6.2;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
Qy 1331 tggacccccgccccgaaggttagacgaagctggtttttatccgtaagcccg 1390
Db 306 TGAGCACACCGTCAGCGGACAGCAGCTCGCGCGAGCTGCACATTGCCATTATAACTC 365
Qy 1391 cgacctggccaggggtctctctctgtagacaatatcaacaccgaagcttatcggggtctct 1450
Db 366 AGACCTTTATCTCTGACGCCAGCACTGCCAGCAACAGTCAGAGGCCCTCGCTGCTCTGGC 425
Qy 1451 tatctacttgagctggagccttcagcagccgggtatcaa 1492
Db 426 TGTTCATGTGATGGGCTCCTTCAATCCGTCCTATGACAA 467

RESULT 15
PCT-US95-09145A-14
Sequence 14, Application PC/TUS9509145A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: RICHARDS, MEDLOCK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09145A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: John A. Harie
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35792CIPPCT
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..822
PCT-US95-09145A-14

Query Match 1.1%; Score 30.8; DB 6; Length 822;
Best Local Similarity 49.4%; Pred. No. 6.2;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
Qy 1331 tggacccccgccccgaaggttagacgaagctggtttttatccgtaagcccg 1390
Db 306 TGAGCACACCGTCAGCGGACAGCAGCTCGCGCGAGCTGCACATTGCCATTATAACTC 365
Qy 1391 cgacctggccaggggtctctctctgtagacaatatcaacaccgaagcttatcggggtctct 1450
Db 366 AGACCTTTATCTCTGACGCCAGCACTGCCAGCAACAGTCAGAGGCCCTCGCTGCTCTGGC 425
Qy 1451 tatctacttgagctggagccttcagcagccgggtatcaa 1492
Db 426 TGTTCATGTGATGGGCTCCTTCAATCCGTCCTATGACAA 467

Search completed: May 17, 2000, 05:51:46
Job time: 4328 sec

Thu May 18 22:21:28 2000

us-09-151-189-1.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:26:14 ; Search time 944.91 Seconds
(without alignments)
12572.623 Million cell updates/sec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcggacaagccttgaaga.....aaatggcgatagattatccc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_est48:*
83: gb_est49:*
84: gb_est50:*
85: gb_est51:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	80.4	2.7	393	81	AW400475	Lamdisest
2	51.6	1.8	925	82	CNS0091P	Drosophil
3	40.4	1.4	942	82	CNS004NS	AL053013 Drosophil
4	38.2	1.3	382	21	T62798	YB99g11.r1
5	38	1.3	324	21	AA365809	AA365809 EST76653
6	37.2	1.3	490	79	AA250569	2821781.5
7	37.2	1.3	504	28	AA062675	zm01f10.s
8	37.2	1.3	752	44	AI314531	AI314531 uJ48b02.y
9	36.8	1.3	468	79	AA284526	LGI.270.C
10	36.8	1.3	1044	83	CNS0150E	AL105704 Drosophil
11	36.4	1.2	533	26	W45709	z27d03.r1
12	36.4	1.2	610	108	AA576692	nbx0089L
13	36.2	1.2	327	36	C74220	C74220 C7420 Rice
14	36	1.2	334	34	AA491660	AA491660 ne7902.s
15	36	1.2	446	50	AA233825	AA233825 u074e09.y
16	36	1.2	557	37	AA171762	AA171762 vt99ell.r
17	36	1.2	600	27	W84020	W84020 mf26908.r1
18	36	1.2	787	82	CNS010B7	AL098749 Drosophil
19	35.8	1.2	488	43	AI205160	AI205160 a084d04.x
20	35.8	1.2	502	79	AA245215	AA245215 2819614.5
21	35.6	1.2	446	25	D83844	D83844.HUMHBC4877
22	35.6	1.2	1069	83	CNS015BE	AL105236 Drosophil
23	35.4	1.2	437	49	AI635496	AI635496 ts65h04.x
24	35.4	1.2	645	82	CNS01213	AL101589 Drosophil
25	35.4	1.2	479	39	AA305715	AA305715 Oj95e12.s
26	35.2	1.2	987	82	CNS0015S	AL074264 Drosophil
27	35.2	1.2	469	70	AW148044	AW148044 da03c09.x
28	35	1.2	1101	82	CNS001XJ	AL078875 Drosophil
29	35	1.2	247	80	AA078869	UI-HF-BLO
30	34.8	1.2	351	32	AA353352	EST61495
31	34.6	1.2	549	103	AO157443	AO157443 nbx0009B
32	34.6	1.2	552	74	AV388992	AV388992 AV388992
33	34.6	1.2	945	82	CNS0076K	AL066880 Drosophil
34	34.6	1.2	592	80	C99982	C99982 C99982 Oryz
35	34.4	1.2	720	49	AI632354	AI632354 tt22h08.x
36	34.4	1.2	775	49	AL048395	AL048395 DRZP586F
37	34.4	1.2	939	82	CNS00CNG	AL059400 Drosophil
38	34.4	1.2	1101	52	CNS00CK2	AL077673 Drosophil
39	34.2	1.2	343	24	N22510	Yw39g12.s1
40	34.2	1.2	384	60	AI800988	AI800988 wg15b08.x
41	34.2	1.2	388	27	AA016068	AA016068 ze32c08.x
42	34.2	1.2	428	44	AI311661	AI311661 q091c06.x
43	34.2	1.2	479	43	AI225585	AI225585 uJ06b01.y
44	34.2	1.2	482	79	AW245544	AW245544 2822724.5

ALIGNMENTS

RESULT	1
LOCUS	AW400475 393 bp mRNA EST 07-FEB-2000
DEFINITION	Lamdisest169est.L digitata sporophyte Lambda ZapII Laminaria digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.
ACCESSION	AW400475
VERSION	AW400475.1 GI:6919074
KEYWORDS	EST.
SOURCE	Laminaria digitata
ORGANISM	Laminaria digitata
REFERENCE	Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group; Phaeophyceae; Laminariales; Laminariaceae; Laminaria.
AUTHORS	Crepineau, F., Roscoe, T., Kaas, R., Kloareg, B. and Boyen, C.
TITLE	Characterisation of complementary DNAs from the expressed Sequence Tag analysis of life cycle stages of Laminaria digitata (Phaeophyceae)
JOURNAL	Unpublished (2000)
COMMENT	On Dec 20, 1995 this sequence version replaced gi:1134111. Contact: Boyen C
	Centre d'Etudes Oceanologique et de Biologie Marine

CNSR-LPR9042, Universite P. & M. Curie BP74, F-29682 Roscoff cedex, France Tel: 33 2 98 29 23 32 Fax: 33 2 98 29 23 24 Email: est@b-roscoff.fr.	Location/Qualifiers 1. 393 /organism="Laminaria digitata" /db_xref="taxon:80365" /clone_lib="L digitata sporophyte Lambda ZapII" /dev_stage="sporophyte" /lab_host="SOLR" /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; The Laminaria sporophyte library, constructed by T Roscoe and F Crepineau, was oligo-(dT) primed and directionally cloned into a Uni-ZAPM XR vector (Stratagene, La Jolla, CA, USA) using total mRNA from sporophytes harvested at 1 file de Sieck (F)."
BASE COUNT	83 a 127 c 107 g 76 t
ORIGIN	
Query Match	2.7%; Score 80.4; DB 81; Length 393;
Best Local Similarity	55.7%; Pred. No. 6e-13;
Matches 215; Conservative	0; Mismatches 166; Indels 5; Gaps 3;
QY 1035	ccctttatgcatgtgacccgacgaataattaccactaccgagccacacccgctgga 1094
DB 13	CCCTCTCGGGAGTTCGCCAAGAACGACTACGTACGCTCGCGCGCAGAGAACCTGCGAGGC 72
QY 1095	atgggagagctcccaatctgacccgctgtcgataggttcggttcggtgacccg 1154
DB 73	CTGCGCGGCTTCAAGGACTCAACATTCCTCCGGAGCGAG--GGAGGAAAGATCGACCT 129
QY 1155	ttctccagctcttcgacgacaccttctgtgttgaacggggccctttctctcag 1214
DB 130	GTCACTGATCTGTTCGACGACCTGCGCGGTGTGACCCGCTCGCTGCTCTCAG 189
QY 1215	ctgtctgtaacagcttaccatgacgtattacggttcgacgacgagagagacatc 1274
DB 190	TTCACTGCTGCGGACTCTCTGATGACTCTATCAAGTCAACCCCGAAGC-TGACCCCTT 248
QY 1275	gcccccactgaactatattgttcgattttgacgaatggtgaacattcagaatgtgga 1334
DB 249	ACCCAGGGGTGGACTACATGACCGCTTCCAGCGCTGCGCTGACGACGACG-GGC 307
QY 1335	ccccccgcccgcgaagtagacgaagctgtgtttatccgttaacgcccgcgac 1394
DB 308	TTCCAAGCTTGAAACTACTTTCGACGAGGAGAACCCGCTTCATCGGACGCGGAGAC 367
QY 1395	ctggccagggctcctctcgtggacaa 1420
DB 368	CTGGCCACTATTCCTTTAGGGATAA 393

RESULT 2

CNS0091P	925 bp DNA GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TE93 end of BAC # BACR19D16 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
DEFINITION	AL053013
ACCESSION	AL053013.1 GI:4934461
VERSION	AL053013.1
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 925)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

FEATURES

source
Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. 334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:910514"
/clone_lib="NCI_CGAP_Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from Ewing's sarcoma.
CDNA made by oligo-dt priming. Non-directionally cloned.
Size selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 86 a 73 c 65 g 109 t 1 others
ORIGIN

Query Match 1.28; Score 36; DB 34; Length 334;
Best Local Similarity 50.3%; Pred. No. 10;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
Qy 1446 tctcttactctagctggagccttcagcagccgggtatcaacgggtccattcattc 1505
Db 149 TCTTCTCCACCCGGGTATGAAGAAATTCATCAGCAGCTGAATATATAAGCCATTCATC 208
Qy 1506 gacagtgatggcagcgggcttcgtcaacttcgacgctctcactctcagattgata 1565
Db 209 CAAAGATGAACAGAGAGACTGTAGTCATCTCTGGGTCTTACTTCTACTGTTG 268
Qy 1566 ggtgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1618
Db 269 GNAGGGCTTCCACTCTCCACGATCTCTCATGTATCAACCAACATGGCCATTGCA 321

RESULT 15

AW323625 446 bp mRNA EST 26-JAN-2000
LOCUS uo74e09.y1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:2648296 5'
DEFINITION similar to gb:D78135 Mouse mRNA for CIRP (MOUSE);, mRNA sequence.
ACCESSION AW323625
VERSION AW323625.1 GI:6757650
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 446)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 22 1998 this sequence version replaced gi:3246932.
Other_ESTs: uo74e09.x1
Contact: Robert Strausberg, Ph.D.
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

MGI:1028748

Seq primer: -40RP from Gibco
High quality sequence stop: 435.
Location/Qualifiers
1. 446
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"

FEATURES

source

/clone="IMAGE:2648296"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: Sali;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 89 a 106 c 164 g 87 t
ORIGIN

Query Match 1.28; Score 36; DB 80; Length 446;
Best Local Similarity 50.6%; Pred. No. 12;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 414 aaggttaagcttccaaacgagcagcagcagcagcagcagcagcagcagcagcagc 473
Db 243 ATGGGAAGTCTGTGGACGGCGCGAGATCAGAGTTGACAGGCTGCAAGTCTTCTGACA 302
Qy 474 agggctccgcatcgtaccgccttggcgtgtgactggtccgcacccctgtttgcatcgcac 533
Db 303 ACCGGTCCCGAGGATACCGGGGTGCTCTGCTGGAGCCGGGGCTTTTCCCTGGGGGAC 362
Qy 534 gaggtcacagcagtggtgttttccctgaggaacacacacacacacacacacacacacac 585
Db 363 GAACCGGGGCCCGAGGGTTCTCCAGAGGAGGAGACCGGGGCTATGGAGG 414

Search completed: May 17, 2000, 05:14:39
JOB time: 2905 sec

OM of: US-09-151-189-2_COPY_441_676 to: GenEmbl.* out_format: pfs
 Date: May 17, 2000 10:56 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL-frame+pin.model -DEV-rlp
 -O/cn2.1/USPTO-spool/US09151189/runat_15052000_082513_1608/app_query.fasta.1
 -DB-GenEmbl -OFMT-fastap -SUFFIX-441-676.rge -GAPOP-12.000
 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
 -GAPOP-4.500 -GAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
 -GAPOP-6.000 -GAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
 -DELOP-6.000 -DELEXT-7.000 -STAR-1 -MATRIX-blossum62
 -TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR SCORE-pct
 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-ext -MINLEN-0
 -MAXLEN-1000000 -USER-US09151189 -NCPU-6 -ICPU-3 -NO_XLPXY -WAT
 -THREADS-1

Search information block:
 Query: US-09-151-189-2_COPY_441_676
 Query length: 236
 Database: GenEmbl.*
 Database sequences: 882769
 Database length: -486395729
 Search time (sec): 483.650000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_p12:AF053411	+	1238.00	2156.07	4.9e-112	2031
gb_p11:DB87657	+	327.50	556.99	5.7e-23	2035
gb_p11:DB87657	+	325.50	554.68	7.7e-23	1794
gb_p11:DB87657	+	318.00	541.49	4.2e-22	1791
gb_p11:DB87657	+	318.00	540.29	4.2e-22	2029
gb_p11:DB87657	+	116.50	168.39	0.2517	12198
gb_p11:DB87657	+	103.50	139.11	10.76	23740
gb_p11:DB87657	+	98.00	132.72	1.88	2103
gb_p11:DB87657	+	95.00	127.86	3.50	2013
gb_p11:DB87657	+	94.00	135.83	11.16	4392
gb_p11:DB87657	+	94.00	135.52	17.06	6055
gb_p11:DB87657	+	92.00	104.72	88.92	105376
gb_p11:DB87657	+	90.50	120.32	119.86	15513
gb_p11:DB87657	+	87.50	122.29	93.13	7297
gb_p11:DB87657	+	87.00	141.42	8.00	909
gb_p11:DB87657	+	87.00	141.25	8.19	926
gb_p11:DB87657	+	87.00	137.82	12.71	1323
gb_p11:DB87657	+	87.00	109.10	505.37	26263
gb_p11:DB87657	+	86.50	140.28	9.26	934
gb_p11:DB87657	+	86.50	121.52	102.80	6583
gb_p11:DB87657	+	86.50	92.57	4.2e+03	133894
gb_p11:DB87657	+	86.50	92.57	4.2e+03	133894
gb_p11:DB87657	+	86.50	87.22	8.4e+03	233554
gb_p11:DB87657	+	86.00	131.98	26.86	2022
gb_p11:DB87657	+	86.00	123.86	77.29	2048
gb_p11:DB87657	+	86.00	123.78	76.95	4704
gb_p11:DB87657	+	86.00	123.77	76.97	4750
gb_p11:DB87657	+	86.00	123.77	76.97	168280
gb_p11:DB87657	+	86.00	89.49	6.3e+03	1830
gb_p11:DB87657	+	85.50	132.06	26.59	1830
gb_p11:DB87657	+	85.50	124.23	72.63	4135
gb_p11:DB87657	+	85.50	114.97	237.96	10830
gb_p11:DB87657	+	85.50	111.66	364.12	15294
gb_p11:DB87657	+	85.00	125.51	61.60	3301
gb_p11:DB87657	+	84.00	94.31	3.4e+03	70680
gb_p11:DB87657	+	84.00	82.96	1.4e+04	230278
gb_p11:DB87657	+	83.50	124.38	71.24	2882
gb_p11:DB87657	+	83.50	83.05	1.4e+04	208124
gb_p11:DB87657	+	83.50	134.62	26.15	887
gb_p11:DB87657	+	82.50	132.20	29.12	1041
gb_p11:DB87657	+	82.50	120.45	117.93	3537

gb_bal:CLOBONT + 82.50 118.80 145.70 4199 L35496 Clostridium botulinu
 gb_bal:CBOBONTFG + 82.50 118.78 146.13 4209 X81714 C.botulinum bont/f g
 gb_bal:U67490 - 82.50 109.38 487.56 11188 U67490 Methanococcus janna
 gb_bal:AE001898 - 82.50 108.30 560.23 12523 AE001898 Deinococcus radio

seq_name: gb_p12:AF053411
 seq documentation block:
 LOCUS AF053411 2931 bp mRNA PLN 12-SEP-1998
 DEFINITION Fucus distichus vanadium bromoperoxidase mRNA, complete cds.
 ACCESSION AF053411
 VERSION AF053411.1 GI:3582762
 KEYWORDS
 SOURCE Fucus distichus.
 ORGANISM Fucus distichus
 Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
 Phaeophyceae; Fucales; Fucaceae; Fucus.
 1 (bases 1 to 2931)
 AUTHORS Vreeland,V., Ng,K. and Epstein,L.
 DIRECT SUBMISSION
 TITLE Submitted (11-MAR-1998) Environmental Science, Policy and
 JOURNAL Management, University of California, 201 Wellman Hall, Berkeley,
 CA 94720-3112, USA
 FEATURES
 Location/Qualifiers
 1..2931
 /organism="Fucus distichus"
 /db_xref="taxon:3012"
 /dev_stage="2-cell embryo"
 1..227
 228..2258
 /codon_start=1
 /product="vanadium bromoperoxidase"
 /protein_id="AAC53279.1"
 /db_xref="GI:3582763"
 /translation="MLCAAATDTGSPMDPTGVLLRLTSEORAKWRRLQEGSKSLGF
 HPESTGYPERKHTGEGRLQPTSDSDALDPTAPNRDNVAFASRRDARRDPT
 EVTASGYPERKHTGEGRLQPTSDSDALDPTAPNRDNVAFASRRDARRDPT
 GVCQITNGEDLATMFKHSLPHDLGOVADDFALDCEILNGDSFCEDEVPAGDPA
 GRVPTAAFAIDISGPAATTPVPTLSSPELAQALAEIYMWALRDVPEMYGT
 TRITTAANLAGMGFPNDVDSIGSDGTVPFQSFQFATFVGTGPFVSQLVNS
 FIDATITVEPQETAFDLYMDFDNLNIONGGPPAGPELDFRIFNARDLAR
 VSFQNTNTEYRSLILLEGALFSGFNGINGPFIQSDQAGFVNGFVSHFRILGAA
 LAQRACQKQVHFARFALGTLHTAGDADFISLLENDELKRVAFINAA
 QPNNEVYLLPQAVGSPSTPSTPGHATONGAFATVVKALIGLDGRCGCPNPVF
 PSDGLLEINFEAGLTYEGEINKLVANVAFGRMLGHIHFYRFDGIGLGLGTTITVT
 LHOELMTAEATFEFLFTGEVTKLFQDGTFSIDGDCSSGLVITGVADCA"
 2259..2931

BASE COUNT 694 a 776 c 784 g 677 t
 ORIGIN
 alignment_scores:
 Quality: 1238.00 Length: 236
 Ratio: 5.246 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-151-189-2_COPY_441_676 x AF053411
 Align seg 1/1 to: AF053411 from: 1 to: 2931

1 HistyPheArgLeuLeuGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
 1548 CACTACTTTCAGATTGATAGTGGCCGAGTGGCGAGCGCTCGCTG 1597
 17 stvrglnysrtrpnlvalhisarphelaarproclualalaLeuGly 34
 1598 TTACCAAGTGGCAGGTGCATGCTGTCACGCCGCCGAGCTCGGG 1647
 34 lyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIle 50
 1648 GTACCTCCACACACACCTTCGGGGGATCTAGATGACAGCTTCGACATC 1697

137 PheGluGluValaCysLeuThrTyrGluGluValaLeuLysLeuAlaVala 153
 1703 TTCAAGGGAAGT...CTCACTGTGGCGTGAATGAACAGCTGCGGA 1749
 153 IasnValaAlaPheGluGluGlnMetLeuGluGluLeuHisTyrArgPheAspG 170
 1750 CAATATTCGATGCGGCGTGAACATGCGAGCTGTTCACTACTCTCTGACC 1799
 170 LylleGlnGluLeuLeuGluGluGluThrValaThrValaThrLeuHis 186
 1800 AGTTGATGCTACTCTCTGCGGAGCAGCTGCGATGGAATCTTGGA 1849
 187 GlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPh 203
 1850 GAGCAAGCTGAGCTATGCGGAGAACTCTTCTTCAACTTGGCGAAGTT 1899
 203 eThrGluGluValleLysLeu 210
 1900 TGATGGAAGTCAATCCAGATC 1921

seq_name: gb_pat:E17199
 seq_documentation_block: 1794 bp DNA PAT 28-JUL-1999
 LOCUS E17199
 DEFINITION Corallina pilulifera mRNA for haloperoxidase.
 ACCESSION E17199
 VERSION E17199.1 GI:5711882
 KEYWORDS JP 1998248581-A/1.
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 1794)
 AUTHORS Izumi, Y. and Tanabe, T.
 TITLE NEW HALOPEXIDASE GENE AND ITS UTILIZATION.
 JOURNAL Patent: JP 1998248581-A 22-SEP-1998;
 OTSUKA PHARMACEUT CO LTD
 OS Corallina pilulifera
 PN JP 1998248581-A/1
 PD 22-SEP-1998
 PF 06-MAR-1997 JP 1997070539
 PI IZUMI YOSHIKAZU, TANABE TADASHI
 PC C12N15/09, C12N9/08, (C12N15/09, C12N1/89), (C12N9/08, C12N1/19);
 CC strandedness: Double;
 CC topology: Linear;
 FH Key
 FH Location/Qualifiers
 FT source 1..1794
 FT /clone='BP01' /organism='Corallina pilulifera' FT
 FT mat_peptide 1..1794
 FT /product='haloperoxidase'.
 FEATURES
 source Location/Qualifiers
 1..1794
 /organism='unidentified'
 /db_xref='taxon:32644'
 BASE COUNT 395 a 481 c 505 g 413 t
 ORIGIN

alignment_scores:
 Quality: 325.50 Length: 224
 Ratio: 2.276 Gaps: 8
 Percent Similarity: 63.839 Percent Identity: 38.393

alignment_block:
 US-09-151-189-2_COPY_441_676 x E17199
 Align seg 1/1 to: E17199 from: 1 to: 1794
 1 HisTyrPheArgLeuLeu...GlyAlaAlaGluLeuAlaGlnArgAlaSe 16
 1138 CACGTCGTGAGTCTGGTCACTGAAGTGGTACGCGCGGTGAAGCGGT 1187

16 rCysTyrGlnLysTyrGlnValHisArgPheAlaArgProGluAlaLeuG 33
 1188 ACGETACCAAGAGTTTAACTATCATCTCGCTGCGCTGCGCTGAGGCTACCG 1237
 33 LylGly...ThrLeuHisAsnThrIleAla...GlyAspLeu 44
 1238 GTGGTCTGATTATTCGTTAACTAAATTCGACCGCAGAGGCGGAGGCAIT 1287
 45 AspaAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysAr 61
 1288 TTCCCTGAGTGTATCTGCTGTGAAGAGCTGGAGATATCTTGGA 1337
 61 gValAlaGluIleAsnAlaAlaGlnAsn...ProAsn 73
 1338 AGCTGAATTAGCAATAGGAAGCAGAACATAGCTGACGGAGATCCTGACC 1387
 73 snGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 89
 1388 CTGATCTCTTCATCTCTGCTGTGGCATTCGCTGCGGAGGCGACCCATTC 1437
 90 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaTh 106
 1438 CATCCGCTCTCGGAAGCGCCACGCTGTGTGCTGGCGCATGTGTGAC 1487
 106 xValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProA 123
 1488 GATCTCGAAGGGCTTC...TTCGACTCCGCGCATCGAGATC...G 1525
 123 snProValPheProSerAsp...AspGlyLeuGluLeuLeuAsn 136
 1526 ATACGTTGTCGAGTTCGACAAAGATGAGGCAAGCTTGTGAAGCTGCT 1575
 137 PheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaVa 153
 1576 TTCAGGGAAGT...CTCACTGTTGCGGCTGAATTAACAAGCTGCGGA 1622
 153 IasnValAlaPheGlyArgGlnMetLeuGluGlyIleHisTyrArgPheAspG 170
 1623 CAATATTCGATGCGGCGCTAAACATGCGAGGTGTCTACTTCTCTGACC 1672
 170 LylleGlnGluLeuLeuGluGluThrIleThrValArgThrLeuHis 186
 1673 AGTTCGAGTCACTCTGCTCGTGGTGGAGGCTGGGATGGATCTTGGA 1722
 187 GluGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPh 203
 1723 GAGCAAGCTGAGTGTGGGAGAACTCTTCTTCACTTGGCGGAGTT 1772
 203 eThrGluGluValleLysLeu 210
 1773 TGATGGAAGTCAATCCAGATC 1794

seq_name: gb_pat:E17200
 seq_documentation_block: 1791 bp DNA PAT 28-JUL-1999
 LOCUS E17200
 DEFINITION Corallina pilulifera mRNA for haloperoxidase.
 ACCESSION E17200
 VERSION E17200.1 GI:5711883
 KEYWORDS JP 1998248581-A/2.
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 1791)
 AUTHORS Izumi, Y. and Tanabe, T.
 TITLE NEW HALOPEXIDASE GENE AND ITS UTILIZATION.
 JOURNAL Patent: JP 1998248581-A 22-SEP-1998;
 OTSUKA PHARMACEUT CO LTD
 OS Corallina pilulifera
 PN JP 1998248581-A/2
 PD 22-SEP-1998
 PF 06-MAR-1997 JP 1997070539

Align seq 1/1 to: D87658 from: 1 to: 2029

```

1 HistyPheArgLeuLeu...GlyAlaGluLeuAlaGlnArgAla 16
1250 CAGCTGCTGAGTCTGCTGAGTGAAGTCTACCGCGCTTGAAGCGGT 1299
16 rCystGlnGlyTyrGlnValHisArgPheAlaArgProGluAlaLeu 33
1300 ACGGTATCAAGAGTTTACATTCATCTCCCTCGCGCTGAGGCTACTG 1349
33 lYcylThrLeu.....HisAsnThrIleAlaGlyAspLeuAsp 45
1350 GTGGTCTGATAGCTTAAACAGAAATCTTTCCTGCGGTAGT..... 1393
46 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 61
1394 .....GATATATATATCCCTGAAGTTAGTGAAGTCTGCTGAGGAGCT 1434
62 .....ValAlaGluLeuAsnAlaAlaGlnAsnProA 72
1435 CTCATCAATCTGATGACGTCTGCTGAGAGCAATGAAACACAGACAGG 1484
72 snAsnGluVal.....ThrTyrLeuLeuProGluAlaIle 83
1485 CTGACGGGATCTGAGCGCGATATATCTTCTGTCGCGATGCAATTT 1534
84 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrG 100
1535 GCGAGGCGAGCCCATTCATCTCCCTCTATGGAAGTGGCCACGCTGTGG 1584
100 nAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly..... 113
1585 TGTGCGCGATGTGTGACAATCTCAAGCGGTCTTCGACGCCAATCTCC 1634
114 .LeuAspArgGlyGlyGlyCysPheProAsnProValPheProSerAsp 129
1635 AGATCATGAAGTGTTCGAG.....GTCGACACTGATGAG 1669
130 AspGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuTyrGluG 146
1570 GACAACCTGTGACACTCTCTTCAAGGAACT...CTCAGTGTGCGG 1716
146 yGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeu 163
1717 TGAATTTGAACAGCTCCGCGCAATGTGCGATCGGCGGACATATGCGG 1766
163 lYleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThr 179
1767 GTGTTCACTACTCTCTGACAGTTCGAGTCACTGCTACTGCTGCTGTAACAG 1816
180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAla 196
1817 ATTGCGATGGAATCTGAGAGCAGCAGCTGACGTATGCGGAGCACTT 1866
196 aThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 210
1867 CTCTTCAATTTGCGGAATTCGACGAACTACATCCAGATT 1909
seq_name: gb_ba2:AE001953
seq documentation block:
LOCUS AE001953 12198 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans R1 section 90 of 229 of the complete chromosome 1.
ACCESSION AE001953 AE000513
VERSION AE001953.1 GI:6458740
KEYWORDS
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans
Bacteria: Thermus/Deinococcus group; Deinococcus.
REFERENCE 1 (bases 1 to 12198)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,

```

```

Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J., Lam, P., McDonald, L.J., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1
Science 286 (5444), 1571-1577 (1999)
20036896
2 (bases 1 to 12198)
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J., Lam, P., McDonald, L.J., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..12198
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
92..997
/gene="DR1011"
92..997
/gene="DR1011"
/feature="DR1011"
/note="similar to GB:AE000657 percent identity: 62.08;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFL0587.1"
/db_xref="GI:6458741"
/transl_table=11
/translation="MPSSAPSSALPDVLHGYSCFNDTFIFHALHAGLVGCLPVO
EYLEDVOTNDNVAAGRLPKFISYRFDVMDYVALRSGGALGCGVGLYVTRGV
KDLALRVASPGALITAEILLALIPGVYVYRMDYVAVGVRGSDYDTADAGLLI
HESRFTPEYGLKLDLGMWEGEGLPLGAILVRDLRDLPAQTORLEQDVARRSLE
YAYAPAAADYIRHAELEMSDEVQAHIDILVNSLSDVGEGERAVRELLURAVEV
GAVTARPELPFLFVGDAAEAGTPTL"
1033..1956
/gene="DR1012"
1033..1956
/gene="DR1012"
/note="similar to GB:AL009126 percent identity: 54.74;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAFL0588.1"
/db_xref="GI:6458742"
/translation="MSKASAPAFPAVEVRGSLKSYGRANVLTDFLNVIPGEVAVTG
PNGAKTILMTLGTAPTECEVRIMGVNVDGPPARALMGAVAEAPAKYPOFTG
TENLKEASLTGAGVGPRISEVLLLELTSMGRRVAEISLQORGLGVAAILAEP
KVLIDEPISGLDPPGIGLIRHIVISLATSGLCAVILSTHRLREIATYAHVIGMTGR
LYIVDURSTISAIYRFRVDDPVGAALLERQPIYVRAIPRAIPPAVAAQLGAQANVPEAL
RLANEDVAVIEVTPDHFIDIEYNERVEYA"
1956..2729
/gene="DR1013"
1956..2729
/gene="DR1013"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFL0594.1"
/db_xref="GI:6458748"
/translation="MLSLISLEFELKLSARIALIVCTLLPFLWSPAPRLGILRV
DLVSGQLPAVSGVQFPLMPLFIAVAVAEIGAEVAGCTGLAPLLRPVSRVIVS
KLIVAFFLLVLTLLGSLVAGIPHGIGFGRGGLGFGVGVLSPSAALAEV

```


TITLE Direct Submission
JOURNAL Submitted (23-JUN-1994) John E. Donelson, Biochemistry, University
of Iowa, 300 Eckstein Medical Research Building, Iowa City, IA
52242, USA

FEATURES

source

1. 2103
/organism="Trypanosoma cruzi"
/db_xref="taxon:5693"
/clone="22"

CDS

237..1889
/product="poly(A) binding protein"
/protein_id="AAC46489.1"
/db_xref="GI:515321"

VTOKSLGIVYVNFONPADAQAGVKGSHIRAKTORPSKRSRGVTVVVK
LPSPDVTALKEMFESKYGLTAIGLATNENGESGYARISEKESAIQAVQVGMV
IDDCATVYERXOPHREELQFNLVYNLDPSYNDKLEKVEFSAFGEVTSKVRDL
GANOTVGFAYAVAHAAKAVEELDEKESPLAKESKLSVCFRSDRERERL
FRERQOQSKYVNLVYNLDPSYNDKLEKVEFSAFGEVTSKVRDL
SEKESAIQAVQVGMVLPSPDVTALKEMFESKYGLTAIGLATNENGESGYARI
LMSGMGVQLGFMGPQAFNGVFPVNRMSMPTPMGCGQLRPMGPTPMNOVRPMA
PQRPMPQIPMAPPOQASQONLSVLANLTPEOKNVLGERLYNHIVAINPAA
AKVTGMLLEMDNGEILNLLDTPGLLDKAVQEALEVLNRHMV

BASE COUNT 563 a 517 c 554 g 469 t
ORIGIN

alignment_scores:

Quality: 98.00 Length: 243
Ratio: 0.933 Gaps: 11
Percent Similarity: 43.210 Percent Identity: 22.634

alignment_block:

US-09-151-189-2_copy_441_676 x TCUI11272/rev ..

Align seq 1/1 to reverse of: TCUI11272 from: 1 to: 2103

1 HistyrPheargLeuileleValaalaGluLeuAlaGlnAargAlaSerCy 17
1370 CATGCTCTCCAGTGTCTGGCAGCAGCATCTTGGGAG..GGCATATG 1324
17 sTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAla..... 31
1323 TCACAAAAGTGG.....ACGCGCCAGGCGCGCTGCTT 1289
32LeuGlyGlyThrLeuHisAsnThrleAlaGlyAspLeu 44
1288 CCATGAAGCTCTGGATGCGACGCGTCCGCGCACTCTG..... 1251
45 AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuIysAr 61
1250CTCTTAAAG 1240
61 gValAlaGluileAsnAlaGlnAsnProAsnAsnGluValThrTyrL 78
1239 A.....ACAAAGCCAAA..... 1227
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyPro 94
1226GCACGGGATACCTAGTGTGCGCG 1203
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuAlaLe 111
1202 GTCCATCAATACATGACAGAAACAGTCTGCCACAGCGCTCAACAGCT 1153
111 ulleGlyLeuAspArgGlyGlyGluCysPheProAsnProValPhePro 128
1152 C..... 1152
128 erAspAspGlyLeuGluLeuileAsnPheGluGlyAlaCysLeuThrTy 144
1151CTTCAGGCGCTCGTGTGACAGATATCTCAAGTGTCTT 1113

145 GluGlyGluileAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMe 161
1112 GACGTAAAGGTTCGGTACTTCTGTG..... 1086
161 tLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyG 178
1085TTGTTGTG..... 1077
178 luthrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe..... 192
1076 ..ACGCTCGCGCGCAGCAGCTTCGCGCTCGCGCTG...ACGTTTCATCTCG 1032
193 ...AlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValI 208
1031 AGAGCGGAAGCGGAGACACTCAGTTTCATCTTCCTTACCCAATGAC 982
208 elysLeuPheGlnAspGlyThrPheSerileAspGlyAspMetCysSerG 225
981 TCTCTTCTCTCGAGTCTCGACAGCTTCGCGCGAGCTTCGTGAGTG 932
225 lylLeuValTyrThrGlyValAlaAspCys 234
931 CGGTACGCGAGCTAGGCAAAATCCGACTGT 903

seq_name: gb_in2.TCU06070

seq_documentation_block:

LOCUS TCU06070 2013 bp mRNA INV 16-DEC-1994
DEFINITION Trypanosoma cruzi Tulahuen poly(A) binding protein mRNA, complete
cgs.

ACCESSION U06070

VERSION U06070.1 GI:459649

KEYWORDS

SOURCE

ORGANISM

Trypanosoma cruzi

Trypanosoma cruzi

Trypanosoma; Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 2013)

REFERENCE

AUTHORS

Batista,J.A., Teixeira,S.M., Donelson,J.E., Kirchhoff,L.V. and de

Sa,C.M.

Characterization of a Trypanosoma cruzi poly(A)-binding protein and

its genes

Mol. Biochem. Parasitol. 67 (2), 301-312 (1994).

JOURNAL

MEDLINE

95174837

REFERENCE

2 (bases 1 to 2013)

AUTHORS

Donelson,J.E.

Direct Submission

Submitted (02-FEB-1994) John E. Donelson, Biochemistry, University

of Iowa, 300 Eckstein Medical Research Bldg., Iowa City, IA 52242,

USA

FEATURES

Location/Qualifiers

1. 2013

/organism="Trypanosoma cruzi"

/strain="Tulahuen"

/db_xref="taxon:5693"

149..1801

/codon_start=1

/product="poly(A) binding protein"

/protein_id="AAC46489.1"

/db_xref="GI:459650"

/translation="MSNFPAAAPSITWGGIDPNLNEQKLYDFVRLGPVASVRVCVDS

LPSPDVTALKEMFESKYGLTAIGLATNENGESGYARISEKESAIQAVQVGMV

IDDCATVYERXOPHREELQFNLVYNLDPSYNDKLEKVEFSAFGEVTSKVRDL

GANOTVGFAYAVAHAAKAVEELDEKESPLAKESKLSVCFRSDRERERL

FRERQOQSKYVNLVYNLDPSYNDKLEKVEFSAFGEVTSKVRDL

SEKESAIQAVQVGMVLPSPDVTALKEMFESKYGLTAIGLATNENGESGYARI

LMSGMGVQLGFMGPQAFNGVFPVNRMSMPTPMGCGQLRPMGPTPMNOVRPMA

PQRPMPQIPMAPPOQASQONLSVLANLTPEOKNVLGERLYNHIVAINPAA

AKVTGMLLEMDNGEILNLLDTPGLLDKAVQEALEVLNRHMV"

BASE COUNT 514 a 497 c 540 g 462 t

ORIGIN

alignment_scores:
 Quality: 95.00 Length: 243
 Ratio: 0.905 Gaps: 11
 Percent Similarity: 43.210 Percent Identity: 22.222

alignment_block:

US-09-151-189-2_COPY_441_676 x TCU06070/rev

Align seg 1/1 to reverse of: TCU06070 from: 1 to: 2013

```

1 HistyrPheArgLeuLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
1282 CATGCTTCACAGTGTCCGACGACGATCTTCGGAG...GGCATATG 1236
17 styrGlnLysTrpGlnValHisArgPheAlaArgProGluAla..... 31
1235 TCACAAAAGTGG.....ACGGCGCCCAAGGCCGCTT 1201
32 .....LeuGlyThrLeuHisAsnThrIleAlaGlyAspLeu 44
1200 CCATGAAGCTCCTGATGCGACGCTCGCGCACTCTG..... 1163
45 AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysAr 61
1162 .....CTCCTTAAAG 1152
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrL 78
1151 A.....ACAAAGCCAA..... 1139
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyPro 94
1138 .....GCACCGGATACCTAGTGGCCG 1115
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 111
1114 GTCCATCATACATGACAGAAACAGTCTCGCACAGCGCTCAACAGCT 1065
111 uileGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPhePros 128
1064 C..... 1064
128 erAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyr 144
1063 .....CTTCAGCGCTCGCTGGTGCAGATATCTCAAGTCTT 1025
145 GluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMe 161
1024 GAGCTAAAGTTCCGGTACTTCGTG..... 998
161 tleuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyG 178
997 .....TTGTGTG..... 989
178 luThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe..... 192
988 ..ACGCTCGCGCGCAGACGTCGCGCTCGCGCTG...ACGTTCACTCG 944
193 ...AlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValI 208
943 AGAGCGAAAGCGCAGACACTTCAGTTTCATCTCTCTTATGCAATGGAC 894
208 eLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerG 225
893 TCCTCTCTCTCGTTCGAGTTCCTCGACACCTTGGCGGACGCTTCGTGAGT 844
225 lyLeuValTyrThrGlyValAlaAspCys. 234
843 GCCTAGCGAACGTAGGCAATCCGACGT 815

```

seq_name: gb_inl:AF044733

seq_documentation_block:

LOCUS AF044733 4292 bp DNA INV 11-FEB-1998
 DEFINITION Trypanosoma cruzi clone Blk poly(A)-binding protein (PABP) gene, complete cds.
 ACCESSION AF044733
 VERSION AF044733.1 GI:2854067
 KEYWORDS
 SOURCE Trypanosoma cruzi
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.
 REFERENCE 1 (bases 1 to 4292)
 AUTHORS Batista,J.A.N. and Martins de Sa,C.
 JOURNAL Direct Submission
 TITLE Submitted (26-JAN-1998) Biologia Celular, Universidade de Brasilia, Campus Universitario - Asa Norte, Brasilia, DF 70910-900, Brazil
 FEATURES
 1. 4292 Location/Qualifiers
 /organism="Trypanosoma cruzi"
 /strain="dm 28c"
 /db_xref="taxon:5693"
 /clone="Blk"
 /note="complement(781..1102)"
 /note="RIME-like element"
 <781..>4292
 /gene="PABP"
 1769..1797
 /gene="PABP"
 /note="trans splicing signal"
 1825..1826
 /gene="PABP"
 /note="secondary spliced leader acceptor site"
 1827..>4292
 /gene="PABP"
 1827..2360
 /gene="PABP"
 1828..1873
 /gene="PABP"
 /note="trans splicing signal"
 1879..1880
 /gene="PABP"
 /note="major spliced leader acceptor site"
 2361..4013
 /gene="PABP"
 /codon_start=1
 /product="poly(A)-binding protein"
 /protein_id="AAC02538.1"
 /db_xref="GI:2854068"
 /translation="MSNPAPASPTWGGGLDPLNEOKLYDHFVRLGPVASVRCVDS
 VTQSLGYGVNFONPDAEAKALDQNGKSHIRIKIQRDPKRSRGVNIYVK
 LPSTVDYALKENRSGKRLPAICLANENCSHGVARISYKESAIQAOVGVNGV
 IDDCALIVRYQHRRERGLQGFNLINLNLDPSYNDERLKEVFSAGVTSKAYRDL
 GANTVGFATVATHEAAKAVEELDEKESPLAKGKMLKVCFRSDRQRERERL
 RRERQOHSKIPNLVKNFDFTVTSERLKEFLERCGETVSVMMDRKPVRSCFGV
 SFKEQSAASRAIOELHGSGTALGPPLFTVYALRKDARROTLEDNRKQPRMCPMG
 LMCGMGPGIOMFGMGPMFNGVFFVNPMSMMPTPMGCGQLRPMGKTPMNOVRAPM
 PORPPMOPINAPPOPSLASQGNLSTVLANTPEQQKNVGLERLYNHIVAINPAA
 AKVTGMLLENDGEILNLDTPGLIDARVQEALEVLNRHNV"
 4014..>4292
 /gene="PABP"
 3'UTR
 BASE COUNT 1183 a 958 c 1056 g 1093 t 2 others
 ORIGIN
 alignment_scores:
 Quality: 94.00 Length: 243
 Ratio: 0.895 Gaps: 11
 Percent Similarity: 43.210 Percent Identity: 22.222
 alignment_block:
 US-09-151-189-2_COPY_441_676 x AF044733/rev

Align seg 1/1 to reverse of: AF044733 from: 1 to: 4292

```

1 HistyrPheargLeuileGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3494 CATGCTCTCCAGTGTCTGGCAGCAGCATCTCTGGCAG...GGCATATG 3448
17 styGlnLysTrpGlnValHisArgPheAlaArgProGluAla..... 31
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3447 TCACAAAAGTGG.....ACGGGCCCAAGGCCGCTGCTT 3413
32 .....LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeu 44
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3412 CCATGAGCTCTGGATGGCAGCGCTCGCCGACCTG..... 3375
45 AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArg 61
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3374 .....CTCTTTAAAG 3364
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAspGluValThrTyrL 78
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3363 A.....ACAAAGCCAA..... 3351
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyPro 94
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3350 .....GCACGGGATACCTAGTGGCCG 3327
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 111
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3326 GTCCATCATAAACACTGACAGAAACAGTCTGCCACAGCGCTCAACAGCT 3277
111 uileGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPheProS 128
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3276 C..... 3276
128 erAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyr 144
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3275 .....CTTCAGCGCTCGCTGGTGACAGTATCGTCAAGAGTCTT. 3237
145 GluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMe 161
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3236 GACGTAAGGTTCCGGTACTGCTGTG..... 3210
161 tLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGly 178
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3209 .....TTGTTGTTG..... 3201
178 luThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe..... 192
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3200 ..AGCTCGCGCGCAGAGCTCGCTCGCTCGCTG...ACGTTCACTCG 3156
193 ...AlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValI 208
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3155 AGACCGAGCGGCAGACACTGATCTATCTCTCTCTAGCAATGAC 3106
208 eLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerg 225
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3105 TCCTCTCTCGTGGAGTCTCTGACAGCCCTGGCGCAGCTCGTGAGTG 3056
225 lyLeuValTyrThrGlyValAlaAspCys 234
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3055 GCGTAGCGCAGCTAGCAATCCGACTGT 3027

```

seq_name: gb_inl:AF044732

```

seq documentation block:
LOCUS AF044732 6055 bp DNA INV 11-FEB-1998
DEFINITION Trypanosoma cruzi clone A6a poly(A)-binding protein (PABP) gene,
complete cds.
ACCESSION AF044732
VERSION AF044732.1 GI:2854065
KEYWORDS Trypanosoma cruzi.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi

```

```

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
1 (bases 1 to 6055)
AUTHORS Batista,J.A.N. and Martins de Sa.C.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-1998) Biologia Celular, Universidade de Brasilia,
Campus Universitario - Asa Norte, Brasilia, DF 70910-900, Brazil
FEATURES
Location/Qualifiers
1..6055
/organism="Trypanosoma cruzi"
/strain="Dm 28c"
/db_xref="taxon:5693"
/cds="A6a"
359..>5270
/gene="PABP"
59..87
/misc_feature
/gene="PABP"
/note="trans splicing signal"
115..116
/misc_feature
/gene="PABP"
/note="secondary spliced leader acceptor site"
117..642
5'UTR
/gene="PABP"
117..5142
mRNA
/gene="PABP"
118..159
/misc_feature
/gene="PABP"
/note="trans splicing signal"
165..166
/misc_feature
/gene="PABP"
/note="major spliced leader acceptor site"
643..2295
CDS
/gene="PABP"
/codon_start=1
/product="poly(A)-binding protein"
/protein_id="AAC02537.1"
/db_xref="gi:2854066"
/translaton="MSNFPAAPSIIVGGLDPLNEQKLYDFVRLGPFVASRVCVDS
VTKSLGYGVNFQNPADAKLDQAGVKLSKRIAKIQDPKSRSGVTNIVVKK
LPSSVDYALKEMESKYGRUTAIGLATNENGESRGYARIYSEKEESAIOAVQVNGMV
IDCAITVETOPHREOLKQFTNLYKNLDPVNDKLEKEVSATGEVTSKRVOL
GANOTGVAVYATVAAAKAVEELDEKESPLAKEMKLSVCRSDRDRERERL
RRERQOHSKYPNLYKNFDPTVTSERLKEPFCGETVSVNMDRATRVSCFGV
SPKQASRAIOELHSGTALGPRLFTVYALRKPDRQTLDEMNKQPMRQPMGVS
LMGMGMPOLGFMGPQAMGVFVNPMSMPTPMGNGOLRPMGTPMNOYARPM
PORPMOPIMAPPPOQSILASGONLSTVLNLTPQOKNVLGERLINHIVAINPAA
AVTGMLEMDNGEILNLLDTPGLLDKQVQEALEVLNRRHNV"
3'UTR
/gene="PABP"
2296..5142
polyA_site
5138
/gene="PABP"
5142
polyA_site
/gene="PABP"
5238..5270
polyA_signal
/gene="PABP"
BASE COUNT 1821 a 1184 c 1468 g 1575 t 7 others
ORIGIN

```

```

alignment_scores:
Quality: 94.00 Length: 243
Ratio: 0.895 Gaps: 11
Percent Similarity: 43.210 Percent Identity: 22.222
alignment_block:
US-09-151-189-2_COPY_441_676 x AF044732/rev
Align seg 1/1 to reverse of: AF044732 from: 1 to: 6055
1 HistyrPheargLeuileGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
1776 CATGCTCTCCAGTGTCTGGCAGCAGCATCTCTGGCAG...GGCATATG.1730

```



```

17  sTyrGlnLysTTPGlnValHisArgPheAlaArgProGluAla..... 31
1729 TCACAAAGTGG.....ACGCGCCCAAGCGCGTGT 1695
32  .....LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeu 44
1694 CCATGAAGCTCTGTGATGCGCGCGCTGCTG..... 1657
45  AspaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 61
1656 .....CTCTTAAAG 1646
61  gValAlaGluIleAsnAlaLeuGlnAsnProAsnGluValThrTyrL 78
1645 A.....ACAAAGCCAA..... 1633
78  euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 94
1632 .....GCACCGGATACCTAGTGGCCG 1609
95  SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 111
1608 GTCCATCATTAACACTGACAAACAGTCTGCCACAGCTCAACACAGCT 1559
111  uileGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProS 128
1558 C..... 1558
128  erAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyr 144
1557 .....CTTCAGCGCTCGCTGTCAGACATGCTCAAGTCTT 1519
145  GluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMe 161
1518 GACGTAAAGTTTCGGTACTGCTGTG..... 1492
161  tLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyG 178
1491 .....TTGTGTGTG..... 1483
178  luThrIleThrValArgThrLeuHisGlnGlnLeuLeuMetThrPhe..... 192
1482 ...ACGCTCGCGCGCAGAGTTCGCGCTCGCGTG...ACGTCATCTCG 1438
193  ...AlaGluAlaThrPheGluPheArgLeuPheThrGlyGluValIle 208
1437 AGACCGAAGCGCGACACACTCAGTTTCATCTCTTAGCAATGGAC 1388
208  eLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerG 225
1387 TCTCTCTCTGCTGAGTCTCTGACAGCTGCGCGCAGCTTCGTGAGTG 1338
225  lyLeuValTyrThrGlyValAlaAspCys 234
1337 GCGTACCGCAGCTGACCAATCCGACTGT 1309

seq_name: gb_p12:YUP8H12
seq_documentation_block:
LOCUS YUP8H12 103576 bp DNA PLN 29-SEP-1997
DEFINITION Arabidopsis thaliana chromosome 1 YAC yup8H12 complete sequence.
ACCESSION AC000098
VERSION AC000098.1 GI:2358139
KEYWORDS HTG.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi.
REFERENCE 1 (bases 1 to 103576)
AUTHORS Theologis,A., Osborne,B.I., Vysotskaya,V.S., Federspiel,N.A.,
Toriumi,M., Yu,G., Ojima,O., Araujo,R., Chung,E., Dewar,K.

```

```

Dietrich,F., Ecker,J.R., Marziani,A., Oefner,P. and Davis,R.W.
The sequence of YAC YUP8H12 from Arabidopsis thaliana chromosome 1
Unpublished (1997)
We have determined that YAC YUP8H12 is chimaeric, and is comprised
of two distinct genomic EcoRI fragments from chromosome 1. This
submission contains the sequence from the EcoRI site at position
181919 to position 285516 (left end) of our previous Phase II
submission. This fragment contains the AXR1 locus and the marker
YUP8H12L.
2 (bases 1 to 103576)
Theologis,A.
Direct Submission
Submitted (25-NOV-1996) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 103576)
Theologis,A., Davis,R.W. and Federspiel,N.A.
Direct Submission (1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 103576)
Theologis,A., Davis,R.W. and Federspiel,N.A.
Direct Submission
Submitted (06-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
5 (bases 1 to 103576)
Theologis,A.
Direct Submission
Submitted (10-SEP-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA
6 (bases 1 to 103576)
Theologis,A.
Direct Submission
Submitted (12-SEP-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA
7 (bases 1 to 103576)
Theologis,A.
Direct Submission
Submitted (29-SEP-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA
On Sep 5, 1997 this sequence version replaced gi:1932823.
FEATURES
source
1. 103576
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="YUP8H12"
3272..6134
/gene="YUP8H12.1"
join(3272..3379,3486..3603,4168..4256,4507..4564,
4922..5057,5173..5236,5466..6134)
/gene="YUP8H12.1"
/codon_start=1
/evidence=not_experimental
/product="YUP8H12.1"
/protein_id="AA571441.1"
/db_xref="GI:2388560"
/translation="MGKREKDHVSVDKVEAVLHLRKHSPLTLKQKFCNRAC
VGRFIRKGNVKAARLSCLSWRSLGIESLIADEFTAEAGLAVAGLDDECR
FVLVAISTMRNVEQVIFLDAIAAALLVTLVLFSLNPLILCNSTAHLPFF
FFFIKYNPKLLQISIFHEHTCTQNSRRGIRTFYDLSTATMIVSMQDFSDYD
DFSSYPSRVSSLRFDTSLSKSTKIGSCASSRAFTVSRDGLDTPKPLCLTDTSS
TKLHGNTGAYISPLNARSFSFASPAARSGPGRFFASTPMPTARTDRHSIGTLR
DPRIPRFSFQSPAFIFRRESHSVSKENPRDSFVQFLKFRFRPYDENTYRSKMRPLG
GLVSTVSTQIRRRHSVLSQRF"
complement(8274..10736)
/gene="YUP8H12.2"
complement(join(8274..8453,8547..8726,8824..8940,
9062..9210,9447..9734,9852..9972,10189..10400,
10691..10736))
/gene="YUP8H12.2"
note="Similar to Arabidopsis hypothetical protein
FID:e326839 (gb129733)".

```

```

/codon_start=1
/evidence-not_experimental
/protein_id="AAB71442.1"
/db_xref="GI:2388561"
/translation="MGYGNRASSKTPALSGREKHODLEKLTLSQPKTLRFVVA
VELVRRSSVLLANVMGLILPSCIFVAPALVITLQPRVAKKIVRESNLFVQV
BELSEYRGLWMLFICVASSGLSGHITFVLGLFHIALFTVIAQCQGLKASAI
YDTQLKRSWLDKCFEFGSFVPSGLPQVQIEALWGLGALGELPFPF
ISRAASGGMKELETCSGDDNGFIARVNOIKSWLSHQINLFTTILLASVNP
LFDLWMCQGFENPWFELALIGLKALIKTHIQTVFIICVNCNOLLDMVENLII
LSFVFGASPELTAKLKMKELIASPPVSSDINVKWDLUSFASVWNGVYVLMLL
NFFGQIVTSAQRYLKKQEEELDALTKNSLTSKSK"
/11547..14661
/gene="YUP8H12.3"
/join(11547..11627,11902..11970,12049..12162,12259..12387,
12472..12540,12819..12974,13147..13183,13265..13347,
13431..13472,13525..13599,13734..13808,13929..14015,
14121..14239,14357..14409,14513..14661)
/gene="YUP8H12.3"
/notes="Similar to Caenorhabditis unknown protein T03F1.1
(gb|U88169)."
/codon_start=1
/evidence-not_experimental
/protein_id="AAB71446.1"
/db_xref="GI:2388585"
/translation="MEVGFALIDLDVLESLSDPALINKLASHVENLALISKCNPH
RSYVLSSEVDSPFISRLMALQRMGIQVNIYERISFVAIVGIGVGSVAEMUT
RCGIGLLLDIDTVELANNRLEFPPQVGMGTDAVOTLAEINPDVYLSFTMNI
RGVGFETFSLTNKSFCPSKREGSGVDVLSCVDNTEARMAVNOACNEQLRTWMS
GVSDAVSGHILQLVPGETACFACAPLWNGGFDVSGIVVAGSIDERTLKREGV
CAELQPTMGVAKPDLVQNSLFLNFGFVSPILGYNLSKDFPVMKRPQCSNVA
CLRSQKIMLAKPERDAKAKMEADASTTIDEGLDHDEWNI SYVDENEKDTTKA
ASSDTPLEGLFEVDAEYKAIATIASGSGTEEDLEDLKQLEALNA"
/15906..16473
/gene="YUP8H12.4"
/join(15906..15936,16035..16176,16428..16473)
/gene="YUP8H12.4"
/notes="EST gb|ATTS0295 comes from this gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAB71443.1"
/db_xref="GI:2388562"
/translation="MSQYDHNQSGANPPPMSTCISPPPIGYPTNQPSHGSVAQK
VETSKDGEFKGCLAMCCGCLDICE"
/complement(15965..17382)
/gene="YUP8H12.5"
/complement(15969..17382)
/gene="YUP8H12.5"
/codon_start=1
/evidence-not_experimental
/product="YUP8H12.5"
/protein_id="AAB71444.1"
/db_xref="GI:2388563"
/translation="MADHNTPPFDLTLDHYIKYOPREAEDEFVHVVKVLGKSS
PLEISFSTYFVEDEDCYELVEYFETEDAGIDAFEAQFLVNDLILVNNTRP
LDEFTGVFKLMAEVLTKPVQLNHAGSOKTESQP"
/complement(17782..20696)
/gene="YUP8H12.6"
/complement(join(17782..17868,17912..18142,18194..18376,
18473..18607,18685..18894,18966..20600,20652..20696))
/gene="YUP8H12.6"
/notes="ESTs gb|AA042402.gb|ATTS1380 come from this gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAB71445.1"
/db_xref="GI:2388564"
/translation="MEETKVSQVPOAKFLKIKKYCDLLAEVTKEDTVMEKEEPT
IFDGGFVKEGINKYDDDDKQKSLDALQHDVYKHLEKLEVEKEDTGLG
LENSKRMKELEDRIKSALEKLELOKQSALEKLSIDERYKTDALLSQ
ALNSVYLOKLSLELSEKLSALIVAEKSKSIQKQVYKSVKLSLSN
QSSARNELEDRLIALQAEHEDIGNVSTKVELQGLFOTLSOLKLEAKELKDL
EAIQVNSLEATLSVAMEKREDENLNAVMEKLSERLEKQARIDEATRTSE
LEALHKHSELKVOKTMEDFSSRDTAKSLTEKSKDLEEKIRVYEGKLAECQGSLSQ

```

```

BELDOSSAENELLADTNOLKIKIOLEGYLDSEKETAIEKLNOKDTEAKDLITKLKS
HENVTEHKROVLBAQSVADTRKVEVEPALLKNTLESTIELEKENGDLAEVNIKLN
OKLANGSETDFOAKLSVLAEKYOAKELQITIEDTLTOLSEREPRLSISLEE
EKNOYELXOSTKELVUKLOLOVDKSKSDMDVSOIEKLSALVAEKSVLESFEQVE
IHLKEVEKVAELSKLOEKHKHAKSDRDVIEKAIQLHKELOASHPAISOKRALSHK
HSELEATLAKSQEELDKAMSVIWHLESKLNLEKQVLAADAKRSVSHIKNHIFIKPVL
QTESITGEEVEKSDLSFSNPKTKIKKNDUASSSGHVSIVKIQAKETWHLMTLK
IAGVALVSVIIGIIICLFCFVVLKIFASTILLEFLIDQARPH"
/23118..24937
/gene="YUP8H12.7"
/join(23118..23497,23582..23753,24134..24340,24432..24691,
24781..24937)
/gene="YUP8H12.7"
/notes="Similar to Prunus pectinesterase (gb|X95991)."
/codon_start=1
/evidence-not_experimental
/protein_id="AAB71446.1"
/db_xref="GI:2388563"
/translation="MKIISISIGIAIIVLASKTLFKTHPEAFGKAIKATSYFKKSL
CDHHHHHHHHHRRKPSDKTKKYSICDDPKNPIDTDTTSLCYLVKNCNCF
TVQSAVDAVGNFSORRVNIWINSIEKVPKPTNITLQSGDFDIAIANDTAYSA
NCTFYCATVOVVGFSQVAKNISPMNVAPIKPGDVGAQVAIRIAGDESFAVCGCFG
AQDTLHDDRGHVEKCYIOGSDIFEGNAKSLYODCHILSMANLSPGSKAVNGAVT
ANGRSSKDSGFSFYNCITIGTGHWLGRWPRYSRVVFTVMTIDVIAPEGNWNFN
DPSRATIFIGEINCSGPGADMSKRAPIVOKLNETOVALLINTSIFDGDOWLOFSL"
/27843..30294
/gene="YUP8H12.8"
/join(27843..28469,29215..29364,29989..30294)
/gene="YUP8H12.8"
/notes="Similar to Arabidopsis Fe(II) transport protein
(gb|U27590)."
/codon_start=1
/evidence-not_experimental
/protein_id="AAB71447.1"
/db_xref="GI:2388566"
/translation="MRITONVKLLFFFFIFSLFIAVSAGESKCEKSHEDDEANKAG
AKKYKIAAPSVLAAGVIGVMPPLGKFFPSLPKPTTFVFFVAFAGVILATGNHV
LPEGYKLTSPCLKGAWPEPTGTIRAMVAAILLSDVSFATSYFKHAKPKTKRIGD
GREDAGSGGGDELGLHVHAGHTHGVGVESGSOVLHRTVRVVAOVLEVLIIVH
SVVIGISGASQSPDPAKALFALPMHQCFFGLGLGCGIAGNFCNMTSITMSIFSV

```

alignment_scores:

	Quality:	92.00	Length:	219
	Ratio:	0.979	Gaps:	9
	Percent Similarity:	42.922	Percent Identity:	21.461
alignment_block:				
US-09-151-189-2_COPY_441_676 x YUP8H12/rev				
Align seg 1/1 to reverse of: YUP8H12 from: 1 to: 103576				
43	AspLeuaspalaasphesaplle.....	SerLeuLeuG1	54	
92433	GATAATGATCCGACAGTATATCCCTGTAGACGACAGAAATCTCTCAA	92384		
54	uasnaspluleuLysargValaalacluleasnaalaalacInAsnP	71		
92383	GAATCGACGATTTTGGATAGATGATTCTTTCAGCTGGAATTCG	92334		
71	roasnasgluValThrTyr.....	LeuLeuProGlnAlaIleGlnVal	85	
92333	CTAACGTTACTTCGCTTACGCTATAACTTTACTCAAGTGGTATCTGT	92284		
86	GlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnG1	102		
92283	GAATTACAGTCAAGAACTTTCTCT.....	GG	92255	
102	YalaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyG	119		
92254	ACTCTCTTCTCCTGATGTTAAAGTCTTCTCTGCTCTCTCGTATGGA	92206		
119	LucYsPheProAsnProValPheProSerAspGlyLeuGluLeuIle	135		

```

92205 ..TTGCTCCGGGAGATGTTATCTTAGCCGTTGCTGACTGAGCTATCT 92158
136 ASN..... 136
|||
92157 AATTCGGGTCGTCGATTCGTTCTTAAAGTTGTTGATGTTGTGAAGGAA 92108
136 ..... 136
92107 TCCGGAGCACATGCTTCTCAGATCGCAAGAGGAGGAGGATTCG 92058
137 ..... 137
92057 AATTCGGGATCAGCCGATGAGGTTTCGTTGGAACG..... 92020
145 GUGUGUGUUAasn...LysLeuAlaValasnValAlaPhe..... 157
|||||:||||:||||:||||| |||
92019 ...GGGAAAATCGGTGTTCACTTATCACCGAATGTTAGATTCGGTAAGGT 91973
158 ..... 158
91972 GAGACGGAAGAACATACCGAAACTTTAGTTTCGCCGCTGAGAAATCT 91923
162 euGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGlu 178
91922 TTGGGCTCTCTCAACT..... 91906
179 ThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluG1 195
91905 .....GTATTCGATGATTTCGAACAGACTTCTTGACTCTCTCAAC 91862
195 uAlaThr 197
|||||
91861 CGCTACT 91855

seq_name: gb_ba2:AE001696
seq_documentation_block:
LOCUS AE001696 15513 bp DNA BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 8 of 136 of the complete genome.
ACCESSION AE001696 AE000512
VERSION AE001696.1 GI:4980582
KEYWORDS
SOURCE Thermotoga maritima.
ORGANISM
REFERENCE
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hart,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
AUTHORS
2 (bases 1 to 15513)
Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hart,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Direct Submission
JOURNAL Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
MEDLINE Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Source
1. 15513
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
complement(75. .458)
/gene="TM0093"
complement(75. .458)

```



```
186 .....HisGlnGluLeuMetThrPheAlaGluGlu 195
9931 ACTTTAAAGACATCATTCGTCATCAACCAACAGCGAACAGACGATC 9882
196 Alatrhrpe.....GluPheArgLeuPheThrGlyGluValIi 208
9881 AGCTTCTTCGCTGCTTCGAGGTTCGATGATGTCGAGCTGCTCGT 9832
208 elysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerG 225
9831 AGATTGATAC.....TTTCGCTTCATTCGAGGTTCCTGCG 9791
225 lyLeuValtyrThr 229
9790 GCCTGCTTATTC 9777
seq_name: gb_v1:ACNPVDNA
seq_documentation_block: 7297 bp DNA VRL 17-FEB-1997
LOCUS ACNPVDNA Autographa californica nuclear polyhedrosis virus DNA.
DEFINITION X71415
ACCESSION X71415
VERSION X71415.1 GI:525306
KEYWORDS glycoprotein gp41; ORF 1137; ORF 141; ORF 1728; ORF 201; ORF 225;
ORF 252; ORF 282; ORF 312; ORF 327; ORF 699; ORF 810; ORF 839.
SOURCE Autographa californica nucleopolyhedrovirus.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 7295)
AUTHORS Zuidema, D.
TITLE Direct Submission
JOURNAL Submitted (14-APR-1993) D. Zuidema, Agricultural University
Wageningen, Dept of Virology, Binnenhaven 11, 6709 PD Wageningen,
NETHERLANDS
REFERENCE 2 (bases 1 to 7297)
AUTHORS Kool, M., Broer, R., Zuidema, D., Goldbach, R.W. and Vlak, J.M.
TITLE Nucleotide sequence and genetic organization of a 7.3 Kb region
(map unit 47 to 52.5) of Autographa californica nuclear
polyhedrosis virus fragment EcoRI-C
J. Gen. Virol. 75 (Pt 3), 487-494 (1994)
94172322
REFERENCE 3 (bases 1 to 7295)
AUTHORS Whitford, M. and Faulkner, P.
TITLE Nucleotide sequence and transcriptional analysis of a gene encoding
gp41, a structural glycoprotein of the baculovirus Autographa
californica nuclear polyhedrosis virus
J. Virol. 66 (8), 4763-4768 (1992)
92333656
JOURNAL Erratum: [[published erratum appears in J Virol 1993
Apr 67(4):2437]]
REMARK On Aug 6, 1994 this sequence version replaced gi:296320.
COMMENT Location/Qualifiers
FEATURES source
1..7297
/organism="Autographa californica nucleopolyhedrovirus"
/strain="ACNPV-E2"
/db_xref="taxon:46015"
/map="47 to 52.5 map units"
/complement(1..339)
/gene="ORF 339"
/complement(<1..339)
/gene="ORF 339"
/codon_start=1
/protein_id="CAA50538.1"
/db_xref="GI:296321"
/db_xref="SWISS-PROT:Q06699"
/translation="MSNLAKEFTEDVSTFTTKVSVVTKLSNWLCRQVTPDKDS
LKLRVYVMEINNEIKLIVETVDSNKLRRQVDFLLALLNNSVFTLLHRF
VDDNVLQDEL"
complement(355..609)
note="ORF 252"
/codon_start=1
CDS
3918
/citation=[3]
/replace="at"
3949
conflict
3972
/citation=[3]
/replace="gc"
3072
conflict
3638..4339
complement(3638..4339)
note="ORF 699"
/codon_start=1
/protein_id="CAA50544.1"
/db_xref="GI:296328"
/db_xref="SWISS-PROT:Q06694"
/translation="MTTTERPPPPPSMSKSTETPLDLSISKSTTETLQRLRNKL
TTLNKYDSELLIYLDHNNKSDYANNINVIKISKVKVKTGASLHAYFAQI
HYSTGYSEPHGPGCPOTFTHTDGLIKLVLIDCECKEEDYIKGNSPNAFR
NESILCRVSFQVLTCTAILLLFNWKFSGMILLILLILSLFCHNNYIISNPY
IEFCNHSKTNKKYR"
3918
/citation=[3]
/replace="at"
3949
conflict
```

```
/protein_id="CAA50539.1"
/db_xref="GI:296322"
/db_xref="SWISS-PROT:Q06690"
/translation="MNYLLGALAIISLVYDKKENSIFLYLLILFLVFLIIVSPAIIS
KNTSEVDIPSHKAKVVRKLEIOTLDSILNKNKTSID"
complement(625..1764)
note="ORF 1137"
/codon_start=1
/protein_id="CAA50540.1"
/db_xref="GI:296323"
/db_xref="SWISS-PROT:Q06687"
/translation="MNGFNVNENNSWKIKQAPRFESVFLATDRORCTPDEVK
NNSLSKMFPPKPTLTKSYKRFKIVYSDVDHLEDSYSLDKEDFSIENOTLL
LIDPELCRRMLERSVTKETQLTINFTNMNLPYKIFPMYMLPTGMDKELNIREK
EKNMLKNVIDTILNFINDIKMLNSDIYVDRGLRGVAFICIMLGTMGRINEROLS
VDDLNLIKRGLHSDTINLKRKRSNNTLNKMKPLELAREIYSRNPILQISKNT
STPFKDFRLLSESGVEMRPSRMIRHYSNLNLYSGVPLQKVAKLMLNHSSASTKH
YLNKINIGLDETSSSEENNDDDDAQHNRNNSGSGESGLYRNE"
complement(1770..2099)
note="ORF 327"
/codon_start=1
/protein_id="CAA50541.1"
/db_xref="GI:296324"
/db_xref="SWISS-PROT:Q06693"
/translation="MNLDPVYRLGNHKEVYIPLKALNDTASEPQOFSEPVHKMP
INDMGYDNTSNVSAGIILISVAFIALELLLVIVYFVILREQQYSDSIDTDSF
FVNKFD"
complement(2102..2416)
note="ORF 312"
/codon_start=1
/protein_id="CAA50542.1"
/db_xref="GI:296326"
/db_xref="SWISS-PROT:Q06692"
/translation="MNTLTNKVWCYVILRODNKLYTGITSNLNRIRIKHKNKOGA
KCLRNTNLRLVHSAYDIKTARMEINLRKRSKIFKRLRIKARPLLDHRLFLAN
KL"
2303..2530
note="ORF 225"
complement(2419..3648)
note="glycoprotein gp41"
/codon_start=1
/protein_id="CAA50543.1"
/db_xref="GI:525307"
/db_xref="SWISS-PROT:P32651"
/translation="MTDERGNFYNTPPPLRVSPNPAIFTSQTYNAPGYVPPATV
PTVATRONMDYTSRSNSTNSVAIAPYKSKETPLDAGESIYWNKCVDFQKIRYV
RCNDSLSPLMILFINTIDMCIDNPISVNVKRFSEETMIRHLIRLOKELGOSN
RAESLSDSNIFQPSFVLSNPAYAKQFYNGADMKGDLAEAKOLSLAVQYMAE
AVTCNPIPLFNOOLANNYMTLLKHATLPNIOQSAVESRPPHINKINDLINAVD
DLFAGGDIYHYLVNENKRVNMSLKENVAFIPLASANIFNFMALATPACQPSM
FONATFLISANVNSPAHLTKSAQESLTELAFQNETLRRFIFQOQINTNRDANAI
AAAPNATNPNTKQRTA"
2618..2620
/citation=[3]
/replace="gc"
3072
conflict
3638..4339
complement(3638..4339)
note="ORF 699"
/codon_start=1
/protein_id="CAA50544.1"
/db_xref="GI:296328"
/db_xref="SWISS-PROT:Q06694"
/translation="MTTTERPPPPPSMSKSTETPLDLSISKSTTETLQRLRNKL
TTLNKYDSELLIYLDHNNKSDYANNINVIKISKVKVKTGASLHAYFAQI
HYSTGYSEPHGPGCPOTFTHTDGLIKLVLIDCECKEEDYIKGNSPNAFR
NESILCRVSFQVLTCTAILLLFNWKFSGMILLILLILSLFCHNNYIISNPY
IEFCNHSKTNKKYR"
3918
/citation=[3]
/replace="at"
3949
conflict
```

```

conflict
/citation={3}
/replace="at"
3962
/citation={3}
/replace="ct"
complement(4188..4730)
/notes="orf 540"
/codon_start=1
/protein_id="CAA50545.1"
/db_xref="GI:525308"
/translation="KASMSNIPDIIVNAGINSEEDENVLDIFIEBYLKLKRALAHI
IKVASFOURLIKYAI SVCGNIGVLCNVGNETDLNMFNCAEIKLKGOMLF
QTKIMSDSKYDAAVHSSPTVETENDDGEASAAAIDBOEGNADVVGLDFEEN
IDGDAPTPKKORLNAQOD"
complement(4586..4729)
/notes="orf 141"
4695..5507
/notes="orf 810"
4696..7239
/notes="orf 2541"
/codon_start=1
/protein_id="CAA50547.1"
/db_xref="GI:1185073"
/db_xref="SWISS-PROT:006670"
/translation="MMSGVLLMLAIFLIIFTLMVLAIFEFDETTKLOVMEY
VKTNADPTPOVIGVSDIMONYIVTWNTVDLSTYHESVHDRIEIPFLAKQO
PVRIYDVRANDPNFEILSGDKADYMKCPATFNEDYALQKCVPPPCDKNSAG
LYPNDLIDTLVLAHQHDKYISNHALYHPIELACFANGHAEVECPDNITDAET
GCKYBELCENRPGTILISIFPSNLLVFMQCVNGHVVGGCPANKIFDRNLMSQVE
AHCATNGAGHTITADIGDQIFKLNNESSQITICINRNSDNQIECSDSRCID
LPNGTGHVFKRVDODISNSGQLVCDNFVYISDIEDQSNVFNALFMDKFRNLNMF
PTEVDTGACVPATADNVNLFRLFAENIPNHYGIDMTMLGTTEVMVKOLVSQDLS
LNDIAFQWLLYARDKDTIGLNFPTGEPIDCFDNLIVDFDARRANICDNGTSVLK
ADLNGGEEFLNVLSTLQKDEYDFQCAISYENGOKIYENEFHFORILTLILQSDYC
TLYTLYQKTYTLSTKYTTPLYNLTKVRPNKIEIYANTRLKNATIPKNAATIP
PVNPFENQPNRQNSILPFLNFQTTDAVYSEFGDDDDHWVYVPTAPPPEPE
PEPEPEPEPEPLPDLNKLDFYCHRYSVFFKLTSCHAENDVIIDALNELNNV
KVADCELDKLDHVLNAYVNGIGRISAYDGDVAVVYKAVSPHYANLQNSND
GVRYNWLHVKNQYMACPEELYNNEFKNIESDKLYTLNLDQSDIV"
complement(4753..4956)
/notes="orf 201"
/codon_start=1
/protein_id="CAA50546.1"
/db_xref="GI:1185072"
/db_xref="SPTREMBL:O64811"
/translation="MHAFFVIGKVDGIEPRYNISVLNIRHVAYVYVCGFVCVGLHI
FSHLELDESREVFKEKINCOIH"
BASE COUNT 1943 a 1607 c 1599 g 2148 t
ORIGIN

```

alignment_scores:
 Quality: 87.50 Length: 252
 Ratio: 0.774 Gaps: 14
 Percent Similarity: 44.841 Percent Identity: 24.206

alignment_block:
 US-09-151-189-2_COPY_441-676 x ACNPVDNA ..

Align seq 1/1 to: ACNPVDNA from: 1 to: 7297

```

16 SerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLe 32
||||| ..... ||||| ..... ||||| ..... ||||| .....
5512 TCCTGCGTGGAA.....GGCATCCGTCGCGCTT 5540

32 uGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheA 49
||||| ..... ||||| ..... ||||| ..... ||||| .....
5541 TACGGCGCGGACACATACATACCGCGGATATCGGCGACACGCAAT 5590

49 spLseSerLeuGluAsnAspGluLeuLeuLysArgValAlaGluLe 65
||||| ..... ||||| ..... ||||| ..... ||||| .....
5591 ATTTCAAATGTTTGAATAAACGAG...TCACAACACTGATACAGTCATC 5637

```

```

66 AsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeu..... 78
||||| ..... ||||| ..... ||||| ..... ||||| .....
5638 AACCGGATCATGAAACTCTGACACAGTAGAGTGTTCGACGACTCCAG 5687

79 .....LeuProGlnAlaIleGlnValGlySerProThrHisPro. 91
||||| ..... ||||| ..... ||||| ..... ||||| .....
5688 ATGCATAGATTATCCCAAC.....GGTACGGGCCACATGATAT 5725

92 .....SerTyrProSerGly..... 96
||||| ..... ||||| ..... ||||| ..... ||||| .....
5726 TCAACACGTTGACGAGCATATTTCGTACACAGCTGGGCCAATGGTGTGC 5775

97 .....HisAlaThrGlnAsnGlyAlaLaph 104
||||| ..... ||||| ..... ||||| ..... ||||| .....
5776 GATAATTTGAAGTATTTCGACATCGAATGTGATCAATCAACGTTGT 5825

104 eAlaThrValLeu.....LysAlaLeuIleGlyLeuAspArgGlyG 118
||||| ..... ||||| ..... ||||| ..... ||||| .....
5826 TGAACACGCTGTTTATGGACAAATTTAGATTAAACATGCAA..... 5868

118 LyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeu 134
||||| ..... ||||| ..... ||||| ..... ||||| .....
5869 .....TTCCCACTGAGGTGTT.....GACGGCACC..... 5895

135 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLe 151
||||| ..... ||||| ..... ||||| ..... ||||| .....
5896 .....GGGTGC.....GTCCAGCCAC 5912

151 uAlaValAsnValAlaPheCylArgGlnMetLeuGlyIle..... 164
||||| ..... ||||| ..... ||||| ..... ||||| .....
5913 CCGGACAATTCACCTTTAGTTTCCACGTTTCCGCTTCCCAAAATTC 5962

165 .....HisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 179
||||| ..... ||||| ..... ||||| ..... ||||| .....
5963 CAACCAATATGTCATCGAC.....ATGCAACACCATCTTGGCAGCAC 6009

180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAl 196
||||| ..... ||||| ..... ||||| ..... ||||| .....
6010 GAAATGGTTAAACAGTTTCCAAAGATTTGCTGTTAAACAACGACGC 6059

196 aThrPheGluPheArgLeu..... 202
||||| ..... ||||| ..... ||||| ..... ||||| .....
6060 CATCTTTCTCAATGGCTTTGTGTCGAGAGACAAAGACACCATCGGC 6109

203 .....PheThrGlyGluValIleLysLeuPheGlnAspGlyThrPhe 216
||||| ..... ||||| ..... ||||| ..... ||||| .....
6110 TTAACCCCTTCCCGCGGAGGATTCGACTGTTTGGAGACACTTGTAC 6159

217 SerIle 218
||||| ..... ||||| ..... ||||| ..... ||||| .....
6160 GATGTG 6165

```

seq_name: gb_pll:AB000134

seq_documentation_block:

LOCUS AB000134 909 bp DNA PLN 05-FEB-1999

DEFINITION Prorocentrum micans mitochondrial DNA for cytochrome oxidase

subunit I, partial cds.

ACCESSION AB000134

VERSION AB000134.1 GI:2897831

KEYWORDS cytochrome oxidase subunit I

SOURCE Prorocentrum micans (strain:NISS12) mitochondrion. DNA.

ORGANISM Mitochondrion Prorocentrum micans

Eukaryota; Alveolata; Dinophyceae; Prorocentrales; Prorocentraceae;

REFERENCE 1 (bases 1 to 909)

AUTHORS Inagaki,Y.

TITLE Direct Submission

JOURNAL Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases. Yuji Inagaki, Biohistry Research Hall, Molecular Evolution Laboratory, Murasaki-cho 1-1, Takatsuki, Osaka 569, Japan (E-mail:yuji.inagaki@ims.brh.co.jp, Tel:0726-81-9761,

OM of: US-09-151-189-2_COPY_441_676 to: N_Geneseq_36:* out_format : pfs

Date: May 17, 2000 11:11 AM

About: Results were produced by the GenCore software; version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame2_b2n_model -DEV=xl
-O=/usr2/1/USPRO_spool/US09151189/runat_15052000_082514_1662/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=441-676.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -START=1 -MATRIX=blomsm2
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsm2
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -MINLEN=0
-MAXLEN=1000000 -USER=US09151189 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:

Query: US-09-151-189-2_COPY_441_676
Query length: 236
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 59.880000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:V56020	1	325.50	689.57	1.1e-29	1794	Haloperoxidase, CP.BP01 encodi
N_Geneseq_36:V56021	1	318.00	653.47	8.6e-29	1791	Haloperoxidase, CP.BP02 encodi
N_Geneseq_36:T03875	1	87.00	152.82	0.6528	2088	Chloroperoxidase cDNA. Use of
N_Geneseq_36:T13683	1	86.50	149.63	0.9831	2520	ACNPV ORF 83, residues 67884-7
N_Geneseq_36:T13683	1	86.50	105.00	301.07	133894	ACNPV genomic DNA clone 6. A
N_Geneseq_36:T08083	1	85.50	151.07	0.8179	1830	Chloroperoxidase gene. Enzymat
N_Geneseq_36:T51612	1	83.50	141.88	2.66	2822	Curvularia verruculosa halope
N_Geneseq_36:T21209.11	1	82.50	98.57	686.76	110000	Continuation (12 of 17) of
N_Geneseq_36:T45009	1	81.50	152.68	0.6653	735	Serratia marcescens IAM 13540 a
N_Geneseq_36:T43062	1	81.50	152.68	0.6653	735	DNA encoding an acid phosphat
N_Geneseq_36:T72100	1	81.50	132.22	9.17	4841	Human RON receptor cDNA. Treat
N_Geneseq_36:T52353	1	80.50	130.77	11.04	4692	Streptococcus pneumoniae genom
N_Geneseq_36:T52179	1	80.50	124.31	25.31	7577	Streptococcus pneumoniae genom
N_Geneseq_36:T18867	1	80.50	119.12	49.20	12019	Alcaligenes sp. Pox proteins
N_Geneseq_36:T24666	1	78.50	151.64	0.7603	453	H. Pylori ORF 06cal0808orf2. H
N_Geneseq_36:T24017	1	78.50	146.96	1.39	687	H. Pylori inner membrane prote
N_Geneseq_36:T32070	1	77.50	141.10	2.34	955	Helicobacter polypeptide GHO 3
N_Geneseq_36:T60336	1	77.50	131.47	10.10	2250	TIM 01 mutant coding sequence
N_Geneseq_36:T47765	1	77.00	129.49	13.02	2438	Comamonas acidovorans derived
N_Geneseq_36:T13304	1	77.00	109.41	170.93	14557	Herpesvirus of turkeys genome
N_Geneseq_36:T35874	1	76.50	119.33	47.93	5471	Herpesvirus of turkeys Banhi F
N_Geneseq_36:T20248.07	1	76.50	85.61	3.6e+03	110000	Continuation (8 of 10) of
N_Geneseq_36:T086907	1	75.50	124.26	25.47	2911	ECORV fragment of Vibrio chol
N_Geneseq_36:T06601	1	75.50	124.25	25.48	2912	Vibrio cholerae strain 395 acc
N_Geneseq_36:T04236	1	75.00	126.51	19.07	2163	Pseudomonas SY77-glutaryl-Cef
N_Geneseq_36:T19076	1	75.00	124.13	25.90	2675	Fibroblast growth factor recep
N_Geneseq_36:T19078	1	75.00	124.10	25.98	2681	Keratinocyte growth factor rec
N_Geneseq_36:T54703	1	74.50	112.95	108.59	5171	DNA sequence of pKktdC38MBA
N_Geneseq_36:T91046	1	74.50	131.26	10.38	1170	DNA encoding sarcosine oxidase
N_Geneseq_36:T2398	1	74.00	130.19	11.90	1387	Bacillus sarcosine oxidase ge
N_Geneseq_36:T60335	1	74.00	119.38	47.63	3369	TIM coding sequence. Nuclear t
N_Geneseq_36:T2128	1	74.00	119.38	47.63	3369	Drosophila timeless gene prote
N_Geneseq_36:T60336	1	74.00	116.98	64.77	4170	TIM splice variant coding sequ
N_Geneseq_36:T2129	1	74.00	116.98	64.77	4170	Drosophila timeless gene prote
N_Geneseq_36:T094735	1	74.00	115.10	82.39	4928	Lycopene cyclase gene. Novel 1
N_Geneseq_36:T40205	1	74.00	115.10	82.39	4928	Cyanobacterial lycopene cyclas
N_Geneseq_36:T60333	1	74.00	114.50	88.97	5198	TIM gene. Nuclear translocat
N_Geneseq_36:T2127	1	74.00	114.50	88.97	5198	Timeless gene. New protein enc
N_Geneseq_36:T21209.16	1	74.00	86.13	3.4e+03	64976	Continuation (17 of 17) of
N_Geneseq_36:T54702	1	73.50	110.79	143.24	6571	DNA sequence of pKktdC38MBA
N_Geneseq_36:T063293	1	73.50	103.49	365.49	12588	Sequence encoding mannuronan

N_Geneseq_36:N40180 - 73.00 126.61 18.83 1460 ! Sequence of recombinant CGF
N_Geneseq_36:N20043 - 73.00 126.61 18.83 1460 ! Pre-prorennin-A gene DNA se
N_Geneseq_36:V00292 - 73.00 123.47 28.18 1931 ! Insecticidal protein encodi
N_Geneseq_36:V69792 + 73.00 117.63 59.55 3246 ! Candida albicans Canik1 gen

seq_name: N_Geneseq_36:V56020

seq_documentation_block:

ID V56020 standard; cDNA; 1794 BP.
AC V56020;
DT 09-DEC-1998 (first entry)
DE Haloperoxidase, CP.BP01 encoding cDNA.
KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
KW Bromine; iodine; ds.
OS Corallina pilulifera.
FH Key Location/Qualifiers
FT 1..1794
FT /*tag= a
FT /product= "Haloperoxidase, CP.BP01"
FT /note= "the stop codon is not indicated"

J10248581-A.

PD 22-SEP-1998.

PF 06-MAR-1997; 070539.

PR 06-MAR-1997; JP-070539.

PA (SAKA) OTSUKA PHARM CO LTD.

DR WPI: 98-560733/48.

DR P-PSDB: W80550.

PT New haloperoxidase gene - and corresponding vector, transformed host

PT Cell and method of preparation

PS Claim 2: Pages 13-15; 20pp; Japanese.

CC This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes

CC encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina

CC pilulifera. A host cell transformed with a vector containing the

CC haloperoxidase genes can be used for the recombinant production of the

CC enzyme. The haloperoxidases of the invention are vanadium-dependent and

CC have halide specificity of bromine and iodine.

CC Sequence 1794 BP; 395A; 481 C; 505 G; 413 T;

alignment_scores:

Quality: 325.50 Length: 224

Ratio: 2.276 Gaps: 8

Percent Similarity: 63.839 Percent Identity: 38.393

alignment_block:

US-09-151-189-2_COPY_441_676 x V56020

Align seg 1/1 to: V56020 from: 1 to: 1794

1 HisTyrPheArgLeuIle...GlyAlaAlaGluLeuAlaGlnArgAla 16

1138 CAGCGTGTGACTGGTGACTGAAGTGGCTACGCCGCTGAAGCGGT 1187

16 rCystyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeu 33

1188 ACGGTACCAAGATTAACTTATCTGCTGCCTGCCTGAGCTACCG 1237

33 LysGly...ThrLeuHisAsnThrIleAla.....GlyAspLeu 44

1238 GTGCTGTATTATCTTAACAAATACGACCCAGAGGCGGAGGAGCAT 1287

45 AspAlaAspPheAspIleSerLeuGluAsnAspGluLeuLeuLysAr 61

1288 TTCCCTGAGGTGATCTTGCTTTGAAGAGCTTGGAGATATCTTGGAGAA 1337

61 GValAlaGluLeuAsnAlaAlaGlnAsn.....ProAsnA 73

1338 AGCTGAATACCATAGGAAACAGACATACGTACCGAGATCTGACC 1387

73 snGluValThrTyLeuLeuProGluAlaIleGlnValGlySerProThr 89

1388 CTGATCCTTCATCTCTTCCGATGCATTCGCCGCGGAGGCCCATTC 1437

90 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThr 106
 11332 CATCGTCTCCTACGAGCGCCACGCTGTGGTGTCTGGCGCATGTGTGAC 1487
 106 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProA 123
 1488 GATCCTGAAGCGTTC.....TTGCACTCCGCGCATCGAGTC.....G 1525
 123 snProValPheProSerAsp.....AspGlyLeuGluLeuLeuAsn 136
 1526 ATCAGGTGTTCAGGTCGACAAAGATGAGGACAACTGTGTGAGTGTCT 1575
 137 PheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsnLysLeuAla 153
 1576 TTCAGGGAAGT...CTCAGTGTGCGGTGAATGACAACTGCGCGCA 1622
 153 IasnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAsp 170
 1623 CAATATGCGATCGCGCGTAAACATGCGAGGTGTCTACTACTCTCTGACC 1672
 170 IyileGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHis 186
 1673 AGTTGCGAGTCACTCTGCTGCGTGACGAGGTTCGCAATGGAATCTTGGAA 1722
 187 GlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPh 203
 1723 GAGCAAGTCTGACGTATGCGGAGAACTTCTTCTCAACTTCCGGAAGTT 1772
 203 eThrGlyGluValIleLysLeu 210
 1773 TCATGGAACTACATCCAGATC 1794
 seq_name: N_Geneseq_36.V56021
 seq_documentation_block:
 ID V56021 standard; cDNA; 1791 BP.
 AC V56021;
 DT 09-DEC-1998 (first entry)
 DE Haloperoxidase, CP.BP02 encoding cDNA.
 KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
 KW bromine; iodine; ds.
 OS Corallina pillulifera.
 FT Key Location/Qualifiers
 FT CDS 1..1791
 FT /*tag= a
 FT /*product= "Haloperoxidase, CP.BP02"
 FT /*note= "the stop codon is not indicated"
 J10248581-A.
 PD 22-SEP-1998.
 PF 06-MAR-1997; 070539.
 PR 06-MAR-1997; JP-070539.
 PA (SAKA) OTSUKA PHARM CO LTD.
 DR WPI: 98-560733/48.
 DR P-PSDB; W80551.
 FT New haloperoxidase gene - and corresponding vector, transformed host
 FT cell and method of preparation
 PS Claim 2; Pages 15-18; 20pp; Japanese.
 CC This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
 CC encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
 CC pillulifera. A host cell transformed with a vector containing the
 CC haloperoxidase genes can be used for the recombinant production of the
 CC enzyme. The haloperoxidases of the invention are vanadium-dependent and
 CC have halide specificity of bromine and iodine.
 SQ Sequence 1791 BP; 390 A; 459 C; 507 G; 435 T;

alignment_scores:
 Quality: 318.00 Length: 231
 Ratio: 2.224 Gaps: 8
 Percent Similarity: 61.905 Percent Identity: 38.095
 alignment_block:
 US-09-151-189-2_copy_441_676 x V56021

Align seg 1/1 to: V56021 from: 1 to: 1791
 1 HisTyrPheArgLeuLeu...GlyAlaAlaGluLeuAlaGlnArgAla 16
 1132 CAGCTCTCAGTCTGTGTGAGTGAAGTGGCTACCGCGGCTTGAAGCGGT 1181
 16 CysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeu 33
 1182 ACGGTATCAGAAAGTTTAACTTCATCTCGCTCGCGCTGAGGCTACTG 1231
 33 IyGlyThrLeu.....HisAsnThrIleAlaGlyAspLeuAsp 45
 1232 GTGCTCTGATTATTCCTTAACAAGAAATCTTCTCTTCCGGGTAGT..... 1275
 46 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 61
 1276GATATAATATTCCTGTGAAGTTAGTGAAGTGGTCGAGGAGCT 1316
 62ValAlaGluLeuAsnAlaAlaGlnAsnProA 72
 1317 CTCATCAATCTCGATGACGTGTCTGAGACCAATGAAAAACAGACAGGG 1366
 72 snAsnGluVal.....ThrTyrLeuLeuProGlnAlaIle 83
 1367 CTGACGGGATTCTGAGCCGCGGATAATCATTTCTGTCCGATGCGCATTT 1416
 84 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrG 100
 1417 GCGAGGCGGCCCATTCCTATCCGCTCTATGAGTGGCCACCGCTGTGGT 1466
 100 nAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly..... 113
 1467 TGTGCGCGCATGTGTGACAAATCTCAAGGCGTCTTCGACGCCCACTTCC 1516
 114 ..LeuAspArgGlyGlyCysPheProAsnProValPheProSerAsp 129
 1517 AGATCATGAAGTGTTCGAG.....GTGACACTGATGAG 1551
 130 AspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGlu 146
 1552 GACAAGCTTGTGAAGTGTCTTTCAAGGGAAGT...CTCACTGTTCGCG 1598
 146 yGluLeuAsnLysLeuAlaValAlaPheGlyArgGlnMetLeuG 163
 1599 TGAATTGAACAAGCTCCCGCAATATTTGGATCGCGCGGAACATGCGGG 1648
 163 IyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 179
 1649 GTGTTCACTACTTCTCTGACCACTTGGATCTGCTACTCTGCTGCTGAG 1698
 180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGlu 196
 1699 ATTGGATTGGAATCTTGGAGGAGCAGAGCTGACGTATGCGGAGAACTT 1748
 196 aThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 210
 1749 CTCTCTCAATTTGCGAAATTCAGCGGAACACTACAAATCCAGATT 1791
 seq_name: N_Geneseq_36.T03875
 seq_documentation_block:
 ID T03875 standard; cDNA; 2088 BP.
 AC T03875;
 DT 25-APR-1996 (first entry)
 DE Chloroperoxidase cDNA
 KW Chloroperoxidase; haloperoxidase; antifouling paint; preservative;
 KW halide; chloride; ss.
 OS Curvularia inaequalis.
 FT Key Location/Qualifiers
 FT CDS 7..1229
 FT /*tag= a
 FT PN W09527009-A1

49 spileSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIle 65
 872 ATTTCAATGTTTGAATAAATACGAG...TCACAAGTATAAGTGCATC 918
 66 AsnAlaAlaGlnAsnProAsnAsnGluValThrTyriLeu..... 78
 919 AACCGGATCAGAAACTCTGACACACAGTACGAGTGTCCGGCGACTCCAG 968
 79LeuProGlnAlaIleGlnValGlySerProThrHisPro. 91
 969 ATGCATAGATTATCCCAAC.....GGTACGGGCGCACATGTTAT 1006
 92SertyrProSerGly..... 96
 1007 TCACAACGTTGACGAGCATATTTCTATACACAGTGGCCAAATGGTGTGC 1056
 97HisAlaThrGlnAsnGlyAlaPhe 104
 1057 GATAATTTGAAGTATTTCCGACATCGAATGTATCAATCAACATGTT 1106
 104 eAlaThrValLeu.....LysAlaLeuIleGlyLeuAspArgGlyG 118
 1107 TCAAAACGCGTGTGTTTATGACAAATTTAGATTAAACATGCAA..... 1149
 118 lylGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 134
 1150TTCCCACTGAGGTGTTT.....GACGGCACC..... 1176
 135 IleAsnPheGluGlyAlaCysLeuThrTyriGluGlyGluIleAsnLysLe 151
 1177CGGTGC.....GTGCCGCCAC 1193
 151 uAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIle..... 164
 1194 CCGGACAACTGTTCACTTTTACGTTCCACGTTGGCCATGAAATATTC 1243
 165HisTyraArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 179
 1244 CAACACCATATGCGATCGAC...ATGCAAACTCCATGTTGGGCACGACC 1290
 180 lIeThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAl 196
 1291 GAAATGGTTAAACAGTGTGGTTCACAAAGATTGTCTGTTAAACACGACGC 1340
 196 aThrPheGluPheArgLeu..... 202
 1341 CATCTTTGCTCAATGCTTTTGTATCGGACAGACAAAGACGCCATCGGC 1390
 203PheThrGlyGluValIleLysLeuPheGlnAspGlyThrPhe 216
 1391 TTAACCGGTTACCGGGGACCTATGCTGCTTTTGGAGAACACTTGTAT 1440
 217 SerIle 218
 1441 GATGTG 1446

seq_name: N_Geneseq_36:T13635

seq_documentation_block:

ID T13635 standard; DNA; 133894 BP.

AC T13635;

DT 03-SEP-1996 (first entry)

DE ACNPV genomic DNA clone 6.

KW Autographa californica nuclear polyhedrosis virus clone 6;

KW disruption; non-essential gene; heterologous protein production;

KW expression vector; baculovirus; ss.

OS Autographa californica nuclear polyhedrosis virus clone 6.

PN W03601320-A2.

PD 18-JAN-1996.

PF 30-JUN-1995; J05078.

PR 04-JUL-1994; GB-013420.

PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.

PI Ayres M, Bishop D, Possee R;
 DR WPI; 96-087670/09.
 PT GENBANK; L22858.
 PT Autographa californica nuclear polyhedrosis virus complete genome
 sequence - useful in the prodn. of vectors for enhanced
 PT heterologous protein expression, such as interleukin(s),
 PT interferon(s) and neurotoxin(s)
 PS Disclosure: Page 90-186; 122pp; English.
 CC The complete nucleotide sequence of the genome of clone 6 of the
 CC baculovirus Autographa californica nuclear polyhedrosis virus (ACNPV)
 CC has been determined. The sequence is taken from the Genbank record
 CC L22858. The patent specification claims a polynucleotide selected from
 CC open reading frames (ORFs) 32, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
 CC 62-63, 66, 86-79, 81-87, 91-92, 96-98, 101-103, 106-126, 128-130,
 CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
 CC the patentees. See T13636-731. Expression vectors contg. the complete
 CC genomic sequence of ACNPV, with the exception that at least one non-
 CC essential ORF is disrupted or replaced are useful for the synthesis of
 CC heterologous proteins.
 SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T;

alignment_scores:

Quality: 86.50

Ratio: 0.765

Percent Similarity: 44.841

Length: 252

Gaps: 14

Percent Identity: 24.206

US-09-151-189-2_COPY_441_676 x T13635

Align seg 1/1 to: T13635 from: 1 to: 133894

16 SerCysTyrGlnIleValHisArgPheAlaArgProGluAlaLe 32

|||||

68700 TCTGCGTGGAA..... 68728

32 uGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 49

|||||

68729 TAACGGCCCGGACACACATGATACGCGCATATCGGCGACACGCAAT 68778

49 spileSerLeuLeuGluAsnAspGlyLeuLeuLysArgValAlaGluIle 65

|||||

68779 ATTCAATGTTTGAATAAATACGAG...TCACAATGATAACGTGCATC 68825

66 AsnAlaAlaGlnAsnProAsnAsnGluValThrTyriLeu..... 78

|||||

68826 AACCGGATCAGAAACTCTGACACACAGTACGAGTGTCCGGCGACTCCAG 68875

79LeuProGlnAlaIleGlnValGlySerProThrHisPro. 91

|||||

68876 ATGCATAGATTATCCCAAC.....GGTACGGGCGCACATGTTAT 68913

92SertyrProSerGly..... 96

|||||

68914 TCACAACGTTGACGAGCATATTTCTGACACAGTGGCCCAATGGTGTGC 68963

97HisAlaThrGlnAsnGlyAlaPhe 104

|||||

68964 GATAATTTGAAGTATTTCCGACATCGAATGTATCAATCAACATGTT 69013

104 eAlaThrValLeu.....LysAlaLeuIleGlyLeuAspArgGlyG 118

|||||

69014 TGAACCGGTTGTTTATGGACAAATTTAGATTAAACATGCAA..... 69056

118 lylGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 134

|||||

69057TTCCCACTGAGGTGTTT.....GACGGCACC..... 69083

135 IleAsnPheGluGlyAlaCysLeuThrTyriGluGlyGluIleAsnLysLe 151

|||||

69084GGGTGC.....GTGCCGCCAC 69100

151 uAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIle..... 164

```

69101 CGCGGACATGCAACATTTTACGTTCCAGCTTGGCATTAATATTC 69150
||||| ||||| ||| ||||| ||||| ||||| ||||| |||||
165 .....HISTYRARGPHEASPGYIIEGNGLYLEULeuLeuGluThr 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69151 CAAACCATTTGGCATGAC...ATGCAACCTCCATGTTGGCAGCAC 69197
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 ILETHRVALARGTHRLEUHisGluGluLeuMetThrPheAlaGluGluAl 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69198 GAAATGTTAAACAGCTGTTTCCAAAGATTGCTGTATACACGACGC 69247
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 aThrPheGluPheArgLeu..... 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69248 CATCTTTGCTCAATGGCTTTGTATGCGAGACAAAGACGCAATCGGGC 69297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
203 .....PHEHRCGLYGLUVALILELysLeuPheGlnAspGlyThrPhe 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69298 TTAACCGGTTACCGCGGAGGCTATACGACTGTTTGGAGACACTTGTAC 69347
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 SerIle 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69348 GATGTG 69353
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: N_Geneseq_36:T00803

seq_documentation_block:

ID T00803 standard; DNA; 1830 BP.

AC T00803;

DT 27-APR-1996 (first entry)

DE Chloroperoxidase gene.

KW Chloroperoxidase; vanadium haloperoxidase; disinfectant;

OS antimicrobial; antiseptic; ss.

PN Curvularia inaequalis strain CBS102.42.

PW W09527046A.

PP 12-OCT-1995; E01229.

PR 31-MAR-1995; E01229.

PA (UNIL) UNILEVER NV.

PI Barnett P, Hondmann DH, Simons LH, Ter Steeg PF;

PI Wever R;

PI WPI: 95-358625/46.

DR Enzymatic antimicrobial compsn, esp. disinfectant - contg. vanadium

PT halo:peroxidase, esp. from Curvularia inaequalis, halide and

PT hydrogen peroxide source

PS Claim 11; Fig 2: 48pp; English.

CC A gene (T00803) encoding the vanadium chloroperoxidase of Curvularia

CC inaequalis CBS102.42 was isolated from a genomic library using a

CC probe obtd. by amplification of cDNA using primers based on isolated

CC fragments of the gene. The gene can be inserted into a vector and

CC used to produce recombinant chloroperoxidase in host cells, e.g.

CC Saccharomyces cerevisiae. The enzyme is used together with a

CC halide source and H2O2 in antimicrobial compositions active against

CC e.g. Streptococcus faecalis and food spoilage microorganisms.

CC Sequence 1830 BP; 402 A; 584 C; 461 G; 383 T;

SQ

alignment_scores:

Quality: 85.50

Ratio: 1.379

Percent Similarity: 44.928

Percent Identity: 21.739

Length: 138

Gaps: 5

alignment_block:

US-09-151-189-2_COPY_441_676 x T00803

Align seg 1/1 to: T00803 from: 1 to: 1830

65 ILEASNALAALGlnAsnProAsnGluValThrTyLeuLeuProG1 81

1141 CTCGGTCCCGACCTAACACACGACGATTCATC..... 1179

81 nAlaIleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

1180 .....AACCTCCCTTTCCAGCTTACCCATCTGGTCAGC 1213
98 IeThrGlnAsnGlyAlaPheAlaThrValLeuLys..... 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1214 CGACCTTTGGCGGTGCTGTGTTCAAATGGTGGCTCGATACATAACAGCG 1263
109 ..... 109
1264 CGCTAGGTACATGAAGGACGACGACCAACCCGACAACTTCCATCGATAT 1313
110 .....AlaLeuIleGlyLeuAspArgGlyGlyGluCysPheP 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1314 GATGATCTCGAGGAGCTCAACGGCTGAACCGGACCTACGCCAGCCTT 1363
122 roAsnProValPheProSerAspArgGlyLeuGluLeuLeu..... 135
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1364 ATGACCCCGACGCGCCCAATCGAAGACCAACCCGATCTGGTGGCACC 1413
136 .....AsnPheGluGlyAlaCys...LeuThrTyGluGlyGluI 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1414 ATTGTTGGCCACTTCGACTCGGCTGGGAACCTCATGTTCCGAAACGCCAT 1463
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleH 165
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1464 TTCGCGCATC.....TTCCTCGGTGTGCC 1486
165 IeTyArgPheAsp 169
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1487 ACTGGCGTTTCGAT 1500
seq_name: N_Geneseq_36:T51612
seq_documentation_block:
ID T51612 standard; DNA; 2822 BP.
AC T51612;
DT 30-APR-1997 (first entry)
DE Curvularia verruculosa haloperoxidase gene.
KW Haloperoxidase; halide oxidation; halogenation; antimicrobial;
KW disinfectant; ss.
OS Curvularia verruculosa strain CBS 147.63.
PW Key Location/Qualifiers
cgs 477..2279
/*tag= a
/*tag= b, complement (477..496)
/*tag= b, primer aHaP1 binding site"
2293..2311
/*tag= c
/*note= "primer aHaP1 binding site"
W09704102-A1.
06-FEB-1997.
09-JUL-1996; U11458.
14-JUL-1995; US-001194.
21-FEB-1996; US-603534.
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO-NORDISK AS.
PI Berka RM, Cherry J, Fuglsang C, Halkier T, Oxenboll KM;
DR WPI: 97-132641/12.
DR P-PSDB; W12042.
PT Haloperoxidase from Curvularia verruculosa - useful for oxidn. of
PT halide, halogenation and, in presence of hydrogen peroxide and
PT thiocyanate, as antimicrobial
PS Claim 28; Page 33-35; 59pp; English.
CC The haloperoxidase gene (T51612) of Curvularia verruculosa CBS.63
CC codes for an enzyme (W12042) that shows optimum activity at about
CC 60 deg C and pH 5.5, which retains at least 75% activity after
CC incubation for 1 hr at pH 7.0 and 60 deg C in the presence of 0.1%
CC H2O2, and which prefers bromide over chloride as substrate. It was
CC isolated from a genomic library using a PCR amplified partial clone
CC (see also T51613-14) as probe. The 2822 bp fragment in a positive
CC plaque has been deposited (in PHAP4.1, in E. coli DH10B) as NRRL
CC B-21519. The gene can be used to express the haloperoxidase in
CC transformed host cells for use in halide oxidation, cpd.

```

CC halogenation and (with H2O2 and thiocyanate) as an antimicrobial.

SQ Sequence 2822 BP; 651 A; 808 C; 674 G; 689 T;

alignment_scores: Quality: 83.50 Length: 138
Ratio: 1.347 Gaps: 5
Percent Similarity: 44.928 Percent Identity: 21.014

alignment_block:

US-09-151-189-2_COPY_441_676 x T51612

Align seg 1/1 to: T51612 from: 1 to: 2822

```

65 ILeAsnAlaLaGlnAsnProAsnAsnGluValThrThrLeuLeuProG1 81
1608 CTCGGTGGCCCGCAGCTACAAACACAAACGACATACCCCTC.....1646
81 mAlaIleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
1647 .....AGCCTCCTTCCCGCCCTACCCATCCTCCATCAGCCAGC 1680
98 IaThrGlnAsnGlyAlaPheAlaThrValLeuLys.....109
1681 CCACCTTTGGCGGTGCTGTATCCAGATGGTCCGCGCTACTACAACGGG 1730
109 .....109
1731 CCGGTAGCAGCTGGAGGACGACGACGACGACGACGACGACGACGACG 1780
110 .....AlaLeuIleGlyLeuAspArgGlyGlyLeuCysPheP 122
1781 GATGATATCCGAGGAGCTCAACGGCGTGAACCGGACCTGCGCCAGCC 1830
122 tOAsnProValPheProSerAspGlyLeuGluLeuLys.....135
1831 AGGACCGGACTGCCCGCCATCGAGACCAACGACGAGTATGCTCGCAG 1880
136 .....AsnPheGluGlyAlaCys...LeuThrThrGlyGluGlu 148
1881 ATCGTGCGCCACTTTGACTCAGCTGGGAAATGATGTCGAAACGCCAT 1930
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLe 165
1931 TTCGCGCATC.....TTCCTCGGCGGTCC 1953
165 IsTyArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967
seq_name: N_Geneseq_36:V21209_11
seq_documentation_block:
Continuation (12 of 17) of V21209 from base 1100001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

```

```

alignment_scores: Quality: 82.50 Length: 91
Ratio: 2.012 Gaps: 2
Percent Similarity: 45.055 Percent Identity: 25.275
alignment_block:
US-09-151-189-2_COPY_441_676 x V21209_11
Align seg 1/1 to: V21209_11 from: 1 to: 110000
91 ProSerTyProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrVa 107
33250 CCAAGCTTCCCACTGGTCACTACAACTTTAGCATTACATGACCAATC 33299
107 lLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlyCysPheProAsn 124
33300 CTTA.....33303
124 roValPheProSerAspGlyLeuGluLeuLeuAsnPheGluGlyAla 140
33303 .....33303
141 CysLeuThrThrGlyGlyGluLeuLeuAsnLysLeu.....AlaVa 153
33304 ..TTATTTACTCAAAACAACTTGAATATTGTTTAAAGTTGGGCTAT 33350
153 lAsnValAlaPheGlyArgGlnMetLeuGlyLeuIleHisTyArgPheAs 170
33351 AATGTAGCTTATAGTAGCTTATGTTGAGTTGAGTTGAGTTGAGTTG 33400
170 lTleGlnGlyLeuLeuGly 177
33401 TCCTGCTGGAATGATTATGGA 33423
seq_name: N_Geneseq_36:T45009
seq_documentation_block:
ID T45009 standard; DNA; 735 BP.
AC T45009;
DT 13-AUG-1997 (first entry)
DE Serratia marcescens IAM 13540 acid phosphatase DNA.
KW IAM 13540; acid phosphatase; production; nucleoside; 5'-phosphate;
KW ester; condiment; pharmaceutical; intermediate; ds.
OS Serratia marcescens
FH Key Location/Qualifiers
FT cds 1..735
FT /tag= a
FT /product= acid_phosphatase
FN W09637603-Al.
PD 28-NOV-1996.
PF 24-MAY-1996; J01402.
PR 25-MAY-1995; JP-149781.
PR 26-MAR-1996; JP-094680.
PA (AJIN) AJINOMOTO CO INC.
PI Asano Y, Mihara Y, Utagawa T, Yamada H;
DR WPI; 97-021215/02.
DR P-PSDB; W06460.
PT Efficient production of nucleoside 5'-phosphate - by reaction of a
PT nucleoside with a phosphoric acid donor in the presence of an acid
PT phosphatase
PS Example 24; Pages 74-75; 95pp; Japanese.
CC The present sequence encodes the Serratia marcescens IAM 13540
CC acid phosphatase (AP), which can be used to produce a nucleoside
CC 5'-phosphate ester from the corresponding nucleoside when a
CC phosphate donor, e.g. poly-phosphoryl- or carbamyl-phosphoric acid,
CC is reacted in its presence at pH 3.0 to 5.5. The PA can be used for
CC the economic and efficient production of nucleoside 5'-phosphates
CC esters for use as condiments, pharmaceuticals and intermediates for
CC pharmaceuticals.
SQ Sequence 735 BP; 180 A; 215 C; 198 G; 142 T;

```

```

alignment_scores:
  Quality: 81.50      Length: 210
  Ratio: 0.823       Gaps: 10
  Percent Similarity: 47.143   Percent Identity: 23.333

alignment_block:
US-09-151-189-2_copy_441_676 x T45009
Align seg 1/1 to: T45009 from: 1 to: 735

13  GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
198  GAAATAGTGGCAATACCTCGCGTGGCAAGCAGGCTTATGATGACGCC 247
29  oGluAlaLeuGly GlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 45
248  ACGTTCGGGGGACGCGC.....GTTGCGCGCGCATTTTCC 282
46  AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeuLy 60
283  ACGGCTTCGGCCTAGAAATAGCCCAACGGAACGCCGAGCTGTTTAA 332
60  sArgValAlaGluLeuAsn.....AlaAlaGlnAsnPr 71
333  GCTGTGTATGAAATAGCTGGAAGACCGCGCGATTGCGACCGCAGCG 382
71  roAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySer 87
383  CCAAAATACACTATGCGCATTCGCCCTTTCGCTTTTATAACGAAGCG 432
88  ProThrHisPro.....SeryTyrPr 94
433  ACCTCCGACCGGACGAAGAACCCCTGTCGAAGACGGTCTTACC 482
94  oSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaL 111
483  TTCGGCCATACCACTACCGCTGGCGACCGCGCTGTGCTGCTGCTGCTG 532
111  euileGlyLeuAspArgGlyGlyGluCysPheProAsnProValPhePro 127
533  TC.....AACCC..... 540
128  SerAspAspGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrty 144
541  .....GCCAG 545
144  rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnM 161
546  GCAGGGTGAATCTCGCAGCGCGCTATGATATGGCCCAAGCGGGTTA 595
161  eLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGly 177
596  TCTGGGTATCACTGCAAGCGAC...GTGACTGGCGCGCGCTGGCG 642
178  GluThrIleThrValArgThrLeuHisGlnGlnLeuMetThrPheAla... 193
643  CGCTCGCCCATGGTGGCGCTTTCATGCCGAA...CCACCTTCGCGCC 689
194  .....GluGluAlaThrPheGluPhe 200
690  CCAGCTGCAGAAAGCCCAAGACGAATTC 717

seq_name: N_Geneseq_36:V43062
seq_documentation_block:
ID V43062 standard; DNA; 735 BP.
AC V43062:
DT 21-OCT-1998 (first entry)
DE DNA encoding an acid phosphatase enzyme.
KW Acid phosphatase; preparation; nucleoside 5'-phosphate ester;
KW seasoning; intermediate; ds.
OS Serratia ficaria.

```

```

FH Key Location/Qualifiers
FT CDS 1..735
FT /tag= a
PD EP-857788-A2.
PD 12-AUG-1998.
PF 20-NOV-1997; 309365.
PR 18-JUN-1997; JP-161674.
PR 21-NOV-1996; JP-311103.
PA (AJIN ) AJINOMOTO CO INC.
PI Asano Y, Mihara Y, Utagawa T, Yamada H;
PI WPI: 98-416010/36.
PI P-PSDB: W71033.
DR Preparation of nucleoside 5'-phosphates comprises reacting
DR nucleoside with phosphate donor in presence of acid phosphatase -
PT used as seasonings or pharmaceutical intermediates
PT Example 24; Pages 47-48; 83pp; English.
PS The present sequence encodes an acid phosphatase enzyme. The
CC specification describes a method for the preparation of nucleoside
CC 5'-phosphate esters. The method comprises reacting a nucleoside with
CC that has been altered to increase its affinity for the nucleoside and/or
CC to increase its thermal stability, or in the presence of a microorganism
CC for such an acid phosphatase. Nucleoside 5'-phosphates are useful as
CC seasonings or pharmaceuticals or as intermediates for them.
CC note: this sequence appears to be claimed (claim 6), but as the claim
CC refers to amino acid sequences, it is clear that the corresponding
CC protein is being claimed.
SQ Sequence 735 BP; 180 A; 215 C; 198 G; 142 T;

```

```

alignment_scores:
  Quality: 81.50      Length: 210
  Ratio: 0.823       Gaps: 10
  Percent Similarity: 47.143   Percent Identity: 23.333

```

alignment_block:

US-09-151-189-2_COPY_441_676 x V43062

Align seg 1/1 to: V43062 from: 1 to: 735

```

13  GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
198  GAAATAGTGGCAATACCTCGCGTGGCAAGCAGGCTTATGATGACGCC 247
29  oGluAlaLeuGly GlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 45
248  ACGTTCGGGGGACGCGC.....GTTGCGCGCGCATTTTCC 282
46  AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeuLy 60
283  ACGGCTTCGGCCTAGAAATAGCCCAACGGAACGCCGAGCTGTTTAA 332
60  sArgValAlaGluLeuAsn.....AlaAlaGlnAsnPr 71
333  GCTGTGTATGAAATAGCTGGAAGACCGCGCGATTGCGACCGCAGCG 382
71  roAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySer 87
383  CCAAAATACACTATGCGCATTCGCCCTTTCGCTTTTATAACGAAGCG 432
88  ProThrHisPro.....SeryTyrPr 94
433  ACCTCCGACCGGACGAAGAACCCCTGTCGAAGACGGTCTTACC 482
94  oSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaL 111
483  TTCGGCCATACCACTACCGCTGGCGACCGCGCTGTGCTGCTGCTGCTG 532
111  euileGlyLeuAspArgGlyGlyGluCysPheProAsnProValPhePro 127
533  TC.....AACCC..... 540

```

```

128 SerAspGlyLeuGluLeuLeuAsnPhGluGlyAlaCysLeuThr 144
541 .....
144 rGluGlyGluLeuLeuAsnValAlaValAlaPheGlyArgGlnM 161
546 GCAGGTTAAATCTCGAGCGCGCTATGATGGCCAAAGCGGGTTA 595
161 etLeuGlyLeuHisThrArgPheAspGlyLeuGluLeuGly 177
596 TCGCGGTATCATCTGCGCAAGCAGC...GTGACTGCGCGCGCTGCGG 642
178 GluThrLeuValArgThrLeuHisGlnGluLeuMetThrPheAla... 193
643 GCGTCGGCATGCGTGGCGCGTTGTCATGCCGAA...CCACCTTCGCGC 689
194 .....GluGluAlaThrPheGluPhe 200
690 CCAGCTGCAAAAGGCCAAAGACGAATTC 717
seq_name: N_Geneseq_36:V72100

seq_documentation_block:
ID V72100 standard; cDNA; 4541 BP.
AC V72100;
DT 12-APR-1999 (first entry)
DE Human RON receptor cDNA
KW MSP; macrophage stimulating protein; apoptosis; human; treatment;
KW neuroendocrine cell; RON receptor; small cell lung carcinoma; tumour;
KW pathogen infection; thrombocyte production; megakaryocyte maturation;
KW thrombocytopenia; hepatocyte growth; ss.
OS Homo sapiens.
FH Key
FT CDS
FT Location/Qualifiers
   29..4231
   /tag= a
   /product= "RON receptor"
   /note= "MSP receptor"

W09855141-A1.
10-DEC-1998.
PF 04-JUN-1997; US-048594.
PA (BGM) BRIGHAM & WOMENS HOSPITAL.
PI Sunday ME, Willet C;
DR WPI: 99-059877/05.
DR P-PSDB: W82791.
PT Treating tumours derived from neuroendocrine cells with macrophage
PT stimulating protein - or its nucleic acid, also for preventing
PS development of these tumours, specifically small cell lung carcinoma
PS Disclosure; Page 73-80; 100pp; English.
CC This sequence encodes a novel human macrophage stimulating protein, MSP,
CC receptor (the RON receptor) which is used in a method for the
CC prophylactic treatment of a tumour derived from neuroendocrine cells
CC (NEC) by administration of this MSP to a subject at risk, sufficient to
CC induce apoptosis of NEC expressing a RON receptor (the receptor for MSP).
CC The method is especially used to treat or prevent small cell lung
CC carcinoma and apoptosis of RON-expressing cells may be induced in vivo or
CC in vitro. Screening NEC from a subject for susceptibility to MSP-induced
CC apoptosis is used to identify patients who will benefit from treatment
CC with the MSP protein. MSP is already known for treating pathogen
CC infections, for stimulating thrombocyte production and megakaryocyte
CC maturation, (for treating thrombocytopaenia) and for stimulating growth of
CC cells (particularly hepatocytes).
SQ Sequence 4541 BP; 862 A; 1384 C; 1309 G; 986 T;

```

```

alignment_scores:
  Quality: 81.50
  Ratio: 0.886
  Percent Similarity: 47.917
  Percent Identity: 25.000
  Gaps: 10
alignment_block:
us-09-151-189-2_copy_441_676 x V72100

```

```

Align seg 1/1 to: V72100 from: 1 to: 4541
49 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysAlaGVal.....62
1034 GAGCTGAGCATGCGGAGGCGCAGGAGTACTATTGGGCTCTTTGTGAC 1083
63 .....AlaGluLeuLeuAlaAlaGlnAsnProAsnAsnGluValThrTyr. 77
1084 TGCAAGAGTGGTGTCTGCTGCGCGCCGCCCACTCTGCTGCTGTGCTG 1133
78 .....LeuLeuProGlnAlaLeuValGly.....86
1134 TCCCCATTGACCTGTCGACACACTAATTATGATGAGGTTGGAGCGCTGT 1183
87 .....SerProThrHisProSerTyrProSerGlyHisAlaThrGlnAs 101
1184 TGTGAATCCCACTGTCATFCCA.....1204
101 nGlyAlaPheAlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyG 118
1205 .....GGCTCCGCGGAGGCC 1220
118 LyGluCysPheProAsnProValPhe...ProSerAspGlyLeuGlu 133
1221 TCACCTTCCTCCAGTCGCCAGTTTGGCCCAACCCGCTGGCTGGAA 1270
134 LeuLeuAsnPhGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsnLy 150
1271 GCCTCAGCCCAACACAGCTGCCGCCACTC.....CC 1305
150 sLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyrA 167
1306 TCTGCTGGTCAGTAGCAGCTTCTCAGCTGGACCTA.....1342
167 rgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArg 183
1343 ..TTCAATGGCTGTGGGACCACTACAGCTCAGCTGCTATGTATGTGACA 1390
184 ThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPh 200
1391 CGCCTT.....GACAAGCTCAGATGGCACACATGGGCACAAATGGATGG 1434
200 eArgLeuPheThrGlyGluValIleLys.....LeuPheGlnA 213
1435 GCATATCTCGAGTGGAGCTGTCAGTCAGTCACCTAACTACTTGTGTATG 1484
213 spGlyThrPheSerIleAspGlyAsp 221
1485 TGTCCAACTCTCACTG...GGTGAC 1507
seq_name: N_Geneseq_36:V52353
seq_documentation_block:
ID V52353 standard; DNA; 4692 BP.
AC V52353;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:220.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09818931-A2.
DD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-27225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 1230-1232; 1409pp; English.

```


190MetThrPheAlaGluGluAlaThrPheGluPheArgLeuPheTh 204
 10212 GCCTTCATCGCCGCGAGAGAAACGCTTCATCATCGCTTCGTTCCA 10261
 204 r.....GlyGluValIleLys 209
 10262 GCTGCAACTGCGAGGAGTCTGAAG 10289

seq_name: N_Geneseq_36:V24666

seq_documentation_block:

ID V24666 standard; DNA; 453 BP.

AC V24666;

DT 02-JUL-1998 (first entry)

DE H. pylori ORF 06cel08orf2.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT CDS 1..453

FT /*tag= a

PN W09737044-A1.

PD 09-OCT-1997.

PP 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PR (ASTR) ASTRA AB.

PA Alm RA, Smith D;

PI WPI; 97-503122/46.

DR P-PSDB; W55257.

DR Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claims 5,6; Page 172; 1145pp; English.

CC This sequence encodes a H. pylori protein of unspecified function.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 35679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

SQ Sequence 453 BP; 89 A; 83 C; 106 G; 175 T;

alignment_scores:

Quality: 78.50 Length: 135

Ratio: 1.246 Gaps: 7

Percent Similarity: 46.667 Percent Identity: 26.667

alignment_block:

US-09-151-189-2_COPY_441_676 x V24666

Align seg 1/1 to: V24666 from: 1 to: 453

48 PheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaG1 64

176 TTTTCTTTTACATCTTTAGTGAATTCACCTTAAATCCCTTAAGCT 125
 64 uileAsnAlaAlaGlnAsnPro.....AsnAsnGluValThrTyrLeuL 79
 126 TTTAGTGGCGGCCCGCCGCTGTAACCAATGGCAATGGTGTGTTT... 171
 79 euProGluAlaIleGluValcylSerProThrHisPro...SerTyrPro 94
 172GCGCATGGCTTTAGTTTCCT 192
 95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 111
 193 ACGGGGCGATGCTTTAGCTTCAGGCGCTTTTACGGCTCTTTGGCGTTGTT 242
 111 uileGlyLeuAspArgglyGlyGluCysPheProAsnProValPhePro 128
 243 GTTA.....TGCATTCTAAC.....G 259
 128 eAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyr 144
 260 CAACAATCGCATTAATAACTATT..... 282
 145 GluGlyGluIleAsnLysLeu.....AlaValAsnValAlaPheGlyAr 159
 283 ...GGTGTATTATTATTACTTTTTCGATTTTAAATGGCGTATGATAG 329
 159 gGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeu 176
 330 GGTATTATTAGGGGCGCATACCTAGCGATGCTTTTAGGAGGGTTTTAT 379
 176 euGly 177
 380 TAGGG 384

OM of: US-09-151-189-2_COPY_441_676 to: Issued_Patents_NA.* out_format : pfs

Date: May 17, 2000 10:57 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame+ p2n.model -DEV-xlp
-O-/cgn2_1/USPTO.spool/US09151189/runat_15052000_082513_1631/app_query.fasta.1
-DB-Issued_Patents_NA -OPT-fastap -SUFFIX-441-676.rni
-GAPOP-12 000 -GAPEXT-4 000 -MINW-PCR-0.100 -LOOPCL-0.000
-LOOPEXT-0.000 -CGAPOP-4.500 -CGAPEXT-0.050 -XGAPOP-10.000
-VGAPEXT-0.500 -VGAPOP-6.000 -VGAPEXT-7.000 -VGAPOP-10.000
-VGAPEXT-0.500 -DELOP-6.000 -DELEXT-7.000 -START-1
-MATRIX-biosum62 -TRANS-human40.cdi -LIST-45 -DOALIGN-200
-THR-SCORE-pct -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-ext
-MINLEN-0 -MAXLEN-1000000 -USER-US09151189 -NCPU-6 -ICPU-3
-NO_XLPHY_WAIT_THREADS-1

Search information block:

Query: US-09-151-189-2_COPY_441_676

Query length: 236

Database: Issued_Patents_NA.*

Database sequences: 226296

Database length: 63486255

Search time (sec): 41.480000

score_list:	Strd Orig	ZScore	Escore Len	Documentation
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-679-405-1 +		83.50	146.22	1.10
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-842-799-1 +		83.50	146.22	1.10
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US96-11458-1 +		83.50	146.22	1.10
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US96-11458-1 +		83.50	146.22	1.10
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-750-145A-23 +		81.50	157.82	0.2486
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-624-601-3 +		81.50	157.82	0.2486
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-931-943-1 +		75.50	127.74	11.79
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-624-601-1 +		75.50	127.74	11.79
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-070-1659-5 +		75.00	127.62	11.96
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-885-418-5 +		75.00	127.62	11.96
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-070-1659-9 +		75.00	127.59	12.00
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-885-418-9 +		75.00	127.59	12.00
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-198-2 +		74.00	122.59	22.80
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-198-4 +		74.00	122.59	22.80
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-399-561-1 +		74.00	118.03	40.91
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-198-8 +		74.00	117.41	44.33
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-198-1 +		74.00	117.39	44.41
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-387-942C-1 +		73.50	105.66	200.07
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-350-492-2 +		73.00	127.00	12.95
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-687-379-5 +		71.50	118.51	38.50
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-687-379-7 +		71.50	118.51	38.50
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-790-912-1 +		71.50	109.62	120.37
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-731-157A-1 +		71.00	121.11	27.55
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-541-780-1 +		71.00	121.11	27.55
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-718-575-3 +		70.50	114.20	66.92
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-481-206-3 +		70.50	114.20	66.92
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-486-269A-3 +		70.50	114.20	66.92
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-221-750A-8 +		69.50	129.36	9.57
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-387-156-9 +		69.50	121.07	27.70
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-694-865-9 +		69.50	121.07	27.70
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-878-748-9 +		69.50	121.07	27.70
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-09-124-491-9 +		69.50	121.07	27.70
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-09-075-087-1 +		69.50	119.86	32.36
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-694-865-15 +		69.50	118.06	40.76
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-09-124-491-15 +		69.50	118.06	40.76
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-455-559-9 +		69.50	116.83	47.72
/cgn2_6/ptodata/1/ina/5C_COMB.seq:PCT-US94-00657-9 +		69.50	116.83	47.72
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-960-332-1 +		69.50	114.65	63.13
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-908-253-1 +		69.50	114.65	63.13
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-455-570A-1 +		69.50	114.65	63.13
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-387-156-5 +		69.50	114.65	63.13

/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-694-865-5 + 69.50 114.65 63.13
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-878-748-5 + 69.50 114.65 63.13
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-535-837-1 + 69.50 114.65 63.13
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-124-491-5 + 69.50 114.65 63.13
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-679-405-1 + 69.50 114.65 63.13

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-679-405-1

seq_documentation_block:

Sequence 1, Application US/08679405

Patent No. 5866393

GENERAL INFORMATION:

APPLICANT: Fuglsang, Claus

APPLICANT: Halkier, Torben

APPLICANT: Oxenboll, Karen M.

APPLICANT: Berka, Randy M.

APPLICANT: Cherry, Joel

TITLE OF INVENTION: Haloperoxidases from Curvularia

TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 58663930 No. 58663930disk of No. 58663930th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/08/679,405

APPLICATION NUMBER: 435

FILING DATE: July 9, 1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/001,194

FILING DATE: July 14, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/603,534

FILING DATE: February 21, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4441.210.US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 867-0123

TELEFAX: (212) 878-9655

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2822 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 477..2276

US-08-679-405-1

alignment_scores:

Quality: 83.50 Length: 138

Ratio: 1.347 Gaps: 5

Percent Similarity: 44.928 Percent Identity: 21.014

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-679-405-1

Align seg 1/1 to: US-08-679-405-1 from: 1 to: 2822

65 tleasnlalalaGlnasProsnAsnGluValThrtyrLeuProG1 81

.....:|||||:.....

```

1608 CTCGGTGGCCCGAGCTACAAACAGACATACCCCTTC.....1646
81 nAlaIleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
    |||||.....
1647 .....AAGCCTCTTTCCCGCGCTACCATCTGCCACG 1680

98 lathrGlnAsnGlyAlaPheAlaThrValLeuLys.....109
    |||||.....
1681 CCACCTTGGCGGTCTCTATTCCAGATGTGCGCGCTACTACAAACGGG 1730
109 .....109
1731 CGCGTAGGCACCTGAAGGAGCAGACACACACATTCGCATTGACAT 1780
110 .....AlaLeuIleGlyLeuAspArgGlyGlyGlyCysPheP 122
1781 GATGATATCGGAGGAGCTCAACGGGTGAACCGGACCTGCCCGCCCT 1830
122 roAsnProValPheProSerAspGlyLeuLeuLys.....135
    |||||.....
1831 ACAGCCGAGCTCCCGCTATCGAGATGTCGACCAACAGGATATCGCGACCCG 1880
136 .....AsnPheGlyGlyAlaCys...LeuThrTyrgLuglyLui 148
1881 ATCGTGGCGGCTTGGCTGCTATTCAGATGTGCGCGCTACTACAAACGG 1930
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleH 165
    |||||.....
1931 TTCTCGCATC.....TTCTCGGCGTCC 1953

165 lAsyArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967

seq name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-842-799-1
seq documentation block:
; Sequence 1, Application US/08842799
; Patent No. 5965418
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Halkier, Torben
; APPLICANT: Oxenboll, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59654180 No. 5965418th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/842,799
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,405
; FILING DATE: July 9, 1996
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,534
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4441.210.US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2822 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 477..2276
US-08-842-799-1

alignment_scores:
Quality: 83.50 Length: 138
Ratio: 1.347 Gaps: 5
Percent Similarity: 44.928 Percent Identity: 21.014

alignment_block:
US-09-151-189-2_COPY_441_676 x US-08-842-799-1
Align seg 1/1 to: US-08-842-799-1 from: 1 to: 2822

65 lIleAsnAlaAlaGlnAsnProAsnAsnGlnValThrTyLeuLeuProGl 81
1608 CTCGGTGGCCCGAGCTACAAACAGACATACCCCTTC.....1646
81 nAlaIleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
    |||||.....
1647 .....AAGCCTCTTTCCCGCGCTACCATCTGCCACG 1680

98 lathrGlnAsnGlyAlaPheAlaThrValLeuLys.....109
    |||||.....
1681 CCACCTTGGCGGTCTCTATTCCAGATGTGCGCGCTACTACAAACGGG 1730
109 .....109
1731 CGCGTAGGCACCTGAAGGAGCAGACACACACATTCGCATTGACAT 1780
110 .....AlaLeuIleGlyLeuAspArgGlyGlyGlyCysPheP 122
1781 GATGATATCGGAGGAGCTCAACGGGTGAACCGGACCTGCCCGCCCT 1830
122 roAsnProValPheProSerAspGlyLeuLeuLys.....135
    |||||.....
1831 ACAGCCGAGCTCCCGCTATCGAGATGTCGACCAACAGGATATCGCGACCCG 1880
136 .....AsnPheGlyGlyAlaCys...LeuThrTyrgLuglyLui 148
1881 ATCGTGGCGGCTTGGCTGCTATTCAGATGTGCGCGCTACTACAAACGG 1930
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleH 165
    |||||.....
1931 TTCTCGCATC.....TTCTCGGCGTCC 1953

165 lAsyArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967

seq_name: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:PCT-US96-11458-1
seq documentation block:
; Sequence 1, Application PC/TUS9611458
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Haloperoxidases from Curvularia

```

```

; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11458
; FILING DATE: 9-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,194
; FILING DATE: 14-JUL-1995
; APPLICATION DATA:
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.204-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
; PCT-US96-11458-1

```

```

alignment_scores:
  Quality: 83.50      Length: 138
  Ratio: 1.347        Gaps: 5
  Percent Similarity: 44.928  Percent Identity: 21.014

```

```
alignment_block:

```

```
US-09-151-189-2_COPY_441_676 x PCT-US96-11458-1
Align seg 1/1 to: PCT-US96-11458-1 from: 1 to: 2822

```

```

65 ILASALALAGLASHProAsnGluValThrTyLeuLeuProgl 81
1608 CTCGGTGGCCAGCTACAAACACACACATACCTTC..... 1646
81 nAlaLeGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
1647 .....AAGCCTCTTTCCCGCGCTACCATCTGCCACG 1680
98 laThrGlnAsnGlyAlaPheAlaThrValLeuLys..... 109
1681 CCACCTTTGGCGGTCTGTATTCCAGATGTCGCGCGCTACTACACACGG 1730
109 ..... 109
1731 CGCGTAGGACCTGGAAGCAGCAGACACACATGTCATGACAT 1780
110 .....AlaLeuLeGlyLeuAspArgGlyGlyGluCysPheP 122
1781 GATGATATCCGAGGAGCTCAACGCGGTGAACGCGGACCTGCGCCGCCCT 1830

```

```

122 IOASnProValPheProSerAspGlyLeuGluLeuLe..... 135
1831 ACGACCGGATGCGCCCATCGAAGACCAACAGGATATCGTCCGCCACCGC 1880
136 .....AsnPheGluGlyAlaCys...LeuThrTyGluGlyGluL 148
1881 ATCTGTGCGCCACATTGACTCAGCTGGGAATGATGTTCCGAAACGCCAT 1930
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyL 165
1931 TTCGCGATC.....TTCCTCGGCGTCC 1953
165 IsTyArgPheAsp 169
1954 ACTGCGCGTTCGAT 1967
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-750-145A-23
seq_documentation_block:
; Sequence 23, Application US/08750145A
; Patent No. 6010851
; GENERAL INFORMATION:
; APPLICANT: MIHARA, Yasuhiro
; APPLICANT: UTAGAWA, Takashi
; APPLICANT: YAMADA, Hideoaki
; APPLICANT: ASANO, Yasuhisa
; TITLE OF INVENTION: Method for Producing Nucleoside-5'-
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFF. DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,145A
; FILING DATE: 01-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-149781
; FILING DATE: 05-May-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-094680
; FILING DATE: 26-Mar-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 0010-0830-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; HYPOSENSITIVE: NO
; ORGANISM: Serratia ficaria
; STRAIN: IAM 13540
; FEATURE:
; NAME/KEY: CDS

```

LOCATION: 1.732
US-08-750-145A-23

alignment_scores:
Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-750-145A-23

Align seg 1/1 to: US-08-750-145A-23 from: 1 to: 735

```

13 GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
   ::::::::::::::::::::|::::::::::::|
198 GAAATAGTGGCAATACCTCGCTGGCAAGCAGGCTTATGATGACGCC 247
   ::::::::::::::::::::|::::::::::::|
29 oGluAlaLeuGly.GlyThrLeuHisAsnThrIleAlaGlyAspLeuasp 45
   ::::::::::::::::::::|::::::::::::|
248 ACCTTCCCGGGGACGCGC.....GTTCCCGCGCATTTTCC 282
   ::::::::::::::::::::|::::::::::::|
46 AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeu 60
   ::::::::::::::::::::|::::::::::::|
283 AACGCTTCGGCTGAGTAATAGCCCAACGGAACCGCGGAGCTTTAA 332
   ::::::::::::::::::::|::::::::::::|
60 sArgValAlaGluIleAsn.....AlaAlaGlnAsnPr 71
   ::::::::::::::::::::|::::::::::::|
333 GCTGCTGATGAAATGCGTGAAGACGCGCGGATTGCGCGACCGCGCG 382
   ::::::::::::::::::::|::::::::::::|
71 roAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySer 87
   ::::::::::::::::::::|::::::::::::|
383 CCAAAATCATTATGCGCATTCGCCCTTTGCGTTTATAACGAGCG 432
   ::::::::::::::::::::|::::::::::::|
88 ProThrHisPro.....SerTyrPr 94
   ::::::::::::::::::::|::::::::::::|
433 ACTGCGGACCGGACGAGAACGACCCCTGCGAAGACGGTCTTACCC 482
   ::::::::::::::::::::|::::::::::::|
94 oSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaL 111
   ::::::::::::::::::::|::::::::::::|
483 TTCGGCCATACCAACATCGCTGGCGACCGCGCTGGTGGCTGAA 532
   ::::::::::::::::::::|::::::::::::|
111 eulLeGlyLeuAspArgGlyGlyCysPheProAsnProValPhePro 127
   ::::::::::::::::::::|::::::::::::|
533 TC.....AACCC 540
   ::::::::::::::::::::|::::::::::::|
128 SerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTy 144
   ::::::::::::::::::::|::::::::::::|
541 .....GCCAG 545
   ::::::::::::::::::::|::::::::::::|
144 rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnM 161
   ::::::::::::::::::::|::::::::::::|
546 GCAGGTGAATCTCGACGCGCTATGATGCGCAACCGCGGTTA 595
   ::::::::::::::::::::|::::::::::::|
161 etLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGly 177
   ::::::::::::::::::::|::::::::::::|
596 TCTCGGTTATCACTGGCAAGCGAC...GTGACTGGCGGCGCATGGCG 642
   ::::::::::::::::::::|::::::::::::|
178 GluThrIleThrValArgThrLeuHisGlnIleLeuMetThrPheAla... 193
   ::::::::::::::::::::|::::::::::::|
643 GCGTGGCCATGTTGGCGGCTTTCATGTCGAA...CCACCTTCGCGCG 689
   ::::::::::::::::::::|::::::::::::|
194 .....GluGluAlaThrPheGluPhe 200
   ::::::::::::::::::::|::::::::::::|
690 CCACCTCAAAAGGCCAAGACGATTC 717
   ::::::::::::::::::::|::::::::::::|

```

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq US-08-975-698A-27

seq_documentation_block:
; Sequence 27, Application US/08975698A
; Patent No. 6015697
; GENERAL INFORMATION:
; APPLICANT: MIHARA, YASUHIRO

```

APPLICANT: UTAGAWA, TAKASHI
APPLICANT: YAMADA, HIDEAKI
APPLICANT: ASANO, YASUHIRO
TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEOSIDE-5'-PHOSPHATE
TITLE OF INVENTION: ESTER
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR.
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/975,698A
APPLICATION NUMBER: US/08/975,698A
FILING DATE: 21-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 0010-0885-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Serratia ficaria
STRAIN: IAM 13540
FEATURE:
NAME/KEY: CDS
LOCATION: 1..732
US-08-975-698A-27

```

```

alignment_scores:
Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333
alignment_block:
US-09-151-189-2_COPY_441_676 x US-08-975-698A-27
Align seg 1/1 to: US-08-975-698A-27 from: 1 to: 735
13 GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
   ::::::::::::::::::::|::::::::::::|
198 GAAATAGTGGCAATACCTCGCTGGCAAGCAGGCTTATGATGACGCC 247
   ::::::::::::::::::::|::::::::::::|
29 oGluAlaLeuGly.GlyThrLeuHisAsnThrIleAlaGlyAspLeuasp 45
   ::::::::::::::::::::|::::::::::::|
248 ACCTTCCCGGGGACGCGC.....GTTCCCGCGCATTTTCC 282
   ::::::::::::::::::::|::::::::::::|
46 AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeu 60
   ::::::::::::::::::::|::::::::::::|
283 AACGCTTCGGCTGAGTAATAGCCCAACGGAACCGCGGAGCTTTAA 332
   ::::::::::::::::::::|::::::::::::|
60 sArgValAlaGluIleAsn.....AlaAlaGlnAsnPr 71
   ::::::::::::::::::::|::::::::::::|
333 GCTGCTGATGAAATGCGTGAAGACGCGCGGATTGCGCGACCGCGCG 382
   ::::::::::::::::::::|::::::::::::|

```



```
71 roAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySer 87
   :::::::::::::: :::: ||| ::::
383 CCAAAATCACTATATGCGCATTCGCCCTTTCGGTTTATACGAAGCG 432
   :::::::::::::: :::: ||| ::::
88 ProThrHisPro.....SerTyrPr 94
   ::::|||
433 ACTTGGCGACCGGACGAAGAAGCACCTGTGCAAGAACGGTTCTTACCC 482
   ::::|||
94 oSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaL 111
   :::::::::::::: :::: ||| ::::
483 TTCGCGCATACCAACCGCTGGCGACCGCGCTGTGCTGGCTGAA 532
   :::::::::::::: :::: ||| ::::
111 euileGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPhePro 127
   ::::|||
533 TC.....ANCCCC..... 540
   ::::
128 SerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTy 144
   ::::
541 .....GCCAG 545
   ::::
144 rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnM 161
   :::::::::::::: :::: ||| ::::
546 GCAGGTGAATCTGCGACGCGCTATGATGGCCAAAGCCGGTTA 595
   :::::::::::::: :::: ||| ::::
161 etLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGly 177
   :::::::::::::: :::: ||| ::::
596 TCTGCGTTATCACTGCGCAAGCGAC...GTGACTGCGGCGCGCATGCG 642
   ::::|||
178 GluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAla... 193
   ::::|||
643 GCGTGGCGCATGTCGCGCTTTCGATGCGGAA...CCGACCTTGGCGCG 689
   ::::|||
194 .....GluGluAlaThrPheGluPhe 200
   ::::::::::::::
590 CCAGCTGCAAAAGGCCAAAGACGAATTC 717
```

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-624-601-3

```
seq_documentation_block:
; Sequence 3, Application US/08624601
; Patent No. 5882653
; GENERAL INFORMATION:
; APPLICANT: Levine Dr., James B.
; TITLE OF INVENTION: Vibrio cholerae O1 (CVD111) and non-O1
; TITLE OF INVENTION: (CVD112 and CVD112M) serogroup vaccine strains, methods
; TITLE OF INVENTION: of making same and products thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spencer & Frank
; STREET: 1100 New York Ave. N.W. Suite 300 East
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/624.601
; FILING DATE: 08-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Schneller Dr., John W.
; REGISTRATION NUMBER: 26,031
; REFERENCE/DOCKET NUMBER: BAWC20019P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)414-4000
; TELEFAX: (202)414-4040
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 1188 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Vibrio cholerae
STRAIN: classical 395
IMMEDIATE SOURCE:
CLONE: orfo
US-08-624-601-3

alignment_scores:
  Quality: 75.50      Length: 236
  Ratio: 0.668       Gaps: 14
  Percent Similarity: 47.881  Percent Identity: 23.729

alignment_block:
US-09-151-189-2_copy_441_676 x US-08-624-601-3

Align seg 1/1 to: US-08-624-601-3 from: 1 to: 1188

32 LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPh 48
   :::::::::::::: ||| ::::::::::::::
445 ATGGCTTCGACTCTC...AATCATGTCAGTGTGTCAGGTGCCACTCTCA 491
   :::::::::::::: ||| ::::::::::::::
48 eaSpIleSerLeuLeuGluAsn.....AspGluLeuLeuLysA 61
   :::::::::::::: ||| ::::::::::::::
492 AGATATGGTTCAGCTCAATACGAAGTTTCACGCGGACCGTGTCTGTAAA 541
   :::::::::::::: ||| ::::::::::::::
61 rGValAlaGluIleAsnAlaIleGlnAsnProAsnAsnGluValThrTyr 77
   ::::|||
542 AAGTGAACCAATCAAC....AATCGATTGAATGGCAGATAAATACTAT 585
   :::::::::::::: ||| ::::::::::::::
78 LeuLeuProGlnAlaIle... 83
   ::::|||
586 CTTGAAGAAGTTCGCGCATCGATGATGGATACACACACGGAGGTCAGAAA 635
   ::::|||
84 .....GlnValGlySerProThrHisProSerT 93
   :::::::::::::: ||| ::::::::::::::
636 AGCCAGAGGATGAACCTCTTCACGCTGTGTCTGTTCACACAGATGTTT 685
   :::::::::::::: ||| ::::::::::::::
93 yrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 109
   ::::|||
686 ACCAAGTAGAATGCT.....GTGCTTCGG 711
   :::::::::::::: ||| ::::::::::::::
110 AlaLeuIleGlyLeu.....AspArgGlyGlyGluCysPhe..... 121
   ::::|||
712 GCGATTGATGAGCTTAAAGATTCTACTCGTGGGGTTGCTGTCGCCGTAA 761
   ::::|||
122 .....ProAsnProValPheProSerAspGlyLeuGluLeuI 135
   :::::::::::::: ||| ::::::::::::::
762 CCCAGACCAACCAATCCACGCCACCGCATAGCAGCAGC.....C 802
   :::::::::::::: ||| ::::::::::::::
135 leAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeu 151
   ::::|||
803 CCAATTATACAGGGCGCTTAATACCATCTCTCAAAAAGCTCAATACCTTA 852
   :::::::::::::: ||| ::::::::::::::
152 .....AlaValAsnValAlaPhe.....GlyAr 159
   :::::::::::::: ||| ::::::::::::::
853 GAGACGATTTACAGCACTCGACACCAATGACGCGGCTATCAGGCGG 902
   :::::::::::::: ||| ::::::::::::::
159 gGlnMetLeuGlyIleHisTyrArgPhe..... 168
   :::::::::::::: ||| ::::::::::::::
903 CTGTAGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 952
   :::::::::::::: ||| ::::::::::::::
169 .....AspGlyIleGlnGlyLeu.....LeuLeuGlyGluThrIle 180
   :::::::::::::: ||| ::::::::::::::
953 AGTTAGAAACGGCTCAGCAGAAATTTAAAGCAGATGATCAACGATAAATC 1002
   :::::::::::::: ||| ::::::::::::::
```

```
181 ThrValArgThrLeuHisGln.....GluLe 189
||||| :|||||
1003 ACCGAGTCGCGCTTTCATCAGTTCAAGGCTCGCGCGGTCGCTTCGTT 1052
: :|||||
189 uMetThrPheAlaGluGlu.....AlaThrPheGluPheArgL 202
: :|||||
1053 TTGCTCTATGTCGAGGAGTTTGTTGTTACAACTCTGTTTACTCTCCC 1102
|||||
202 euPheThr 204
|||||
1103 TCTTTTCT 1110

seq_name: /cnp2_6/ptodata/1/ina/5A_COMB.seq:US-07-931-943-1
seq_documentation_block:
; Sequence 1, Application US/07931943
; Patent No. 5470729
; GENERAL INFORMATION:
; APPLICANT: KAPER, James B.
; APPLICANT: BAUDRY-MAURELLI, Bernadette
; APPLICANT: FASANO, Alessio
; TITLE OF INVENTION: METHOD OF ISOLATING RESTRICTION FRAGMENT
; TITLE OF INVENTION: DELETIONS IN VIBRIO CHOLERAE, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & MCKENZIE
; STREET: 815 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/931,943
; FILING DATE: 19920812
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/821,072
; FILING DATE: 16-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/533,315
; FILING DATE: 05-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/581,406
; FILING DATE: 17-FEB-1984
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/472,276
; FILING DATE: 04-MAR-1983
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/363,383
; FILING DATE: 05-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/867,633
; FILING DATE: 27-MAY-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Kile, Bradford E.
; REGISTRATION NUMBER: 25,223
; REFERENCE/DOCKET NUMBER: BAWCZ0016P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 452-7000
; TELEFAX: (202) 452-7074
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2912 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
```

```
ORGANISM: Vibrio cholerae
STRAIN: CLASSICAL 395
FEATURE:
NAME/KEY: CDS
LOCATION: 1034..2218
FEATURE:
NAME/KEY: CDS
LOCATION: 2221..2508
US-07-931-943-1

alignment_scores:
Quality: 75.50 Length: 236
Ratio: 0.668 Gaps: 14
Percent Similarity: 47.881 Percent Identity: 23.729

alignment_block:
US-09-151-189-2_COPY_441_676 x US-07-931-943-1 ..
Align seg 1/1 to: US-07-931-943-1 from: 1 to: 2912
32 LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPh 48
: :|||||
1478 ATGGCTTCGACTCTC...AATCATGTCAGTGTGTCAGGTGCCACTCTCA 1524
|||||
48 eaSpIleSerLeuLeuGluAsn.....AspGluLeuLeuLysA 61
|||||
1525 AGATATGTTTCAGCTCAATACGAAGTTTCACGGCGACCGTCTTGAA 1574
|||||
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTr 77
|||||
1575 AGTGAACGCAATCAAC.....AATCGTTGAATGGCAGATAACTAT 1618
|||||
78 LeuLeuProGlnAlaIle..... 83
|||||
1619 CTTGAAGAAGTTCCGATCATGATATGGGATACACACGGGAGGTCAAAA 1668
84 .....GlnValGlySerProThrHisProSerT 93
: :|||||
1669 AGCCNAGGATGAACCTCTCTTCACGTGTGTTGTTGCACACGATGTT 1718
93 yrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 109
||| :|||||
1719 ACCAAAGTAAGAATGCT.....GTGCTTCGG 1744
110 AlaLeuIleGlyLeu.....AspArgGlyGlyGluCysPhe..... 121
|||||
1745 GCGATTGATGAGCTTAAAGATTCACGTCGTGGGTTGTCGTCGCCCTAA 1794
122 .....ProAsnProValPheProSerAspGlyLeuGluLeuI 135
: :|||||
1795 CCGACACCAACCAATCCACCGCCAGCATAGCAGCAGC.....C 1835
135 leAsnPheGluGlyAlaCysLeuThrGlyGluIleAsnLysLeu 151
|||||
1836 CCAATTATACAGGGGCGCTTAATACCATCTCTAAAAAGCTCAATACCTTA 1885
152 .....AlaValAsnValAlaPhe...GlyAr 159
: :|||||
1886 GAGACGATTTACAGCAACTCGACACCATGAACACGGCGCTATCAGGGCG 1935
159 gGlnMetLeuGlyIleHisTyArgPhe..... 168
1936 CTGTAGTAACCTCGCTCGCTGTCAGTTCCGATACGCGAGCGGAGACCG 1985
169 .....AspGlyIleGlnGlyLeu.....LeuLeuGlyGluThrIle 180
1986 ACTTAGAAGCGCTCAGCAGCAATTTTAACGACGAGATGATCAACAGTAATC 2035
181 ThrValArgThrLeuHisGln.....GluLe 189
|||||
2036 ACCCAGTCGCGCTTTCATCAGTTCAAGGCTCGCGCGGTCGCTTCGTT 2085
```



```

COUNTRY: USA
ZIP: 43210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/070,165F
FILING DATE:
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614)-293-8093
TELEFAX: (614)-293-5631
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2675 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5750365ophthalmus viridescens
DEVELOPMENTAL STAGE: Adult
TISSUE TYPE: Regenerating forelimb blastema
CELL TYPE: Mesenchyme and Epithelium
IMMEDIATE SOURCE:
LIBRARY: lambda gt11
CLONE: KP23-1
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..324
FEATURE:
NAME/KEY: CDS
LOCATION: 325..2511
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 2512..2675
PUBLICATION INFORMATION:
AUTHORS: Poulin, Matthew L
TITLE: Nucleotide sequences of two new
TITLE: (No. 5750365ophthalmus viridescens) fibroblast growth
TITLE: factor receptor-2 variants
JOURNAL: Biochim. Biophys. Acta
VOLUME: 1220
PAGES: 208-211
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 2675
US-08-070-165F-5

alignment_scores:
Quality: 75.00 Length: 202
Ratio: 0.833 Gaps: 10
Percent Similarity: 44.554 Percent Identity: 22.277

alignment_block:
US-09-151-189-2_COPY_441_676 x US-08-070-165F-5
Align seg 1/1 to: US-08-070-165F-5 from: 1 to: 2675
44 LeuAspAlaGluLeuAsnValGlnValGlySerProThrHisPro.....91
1336 GTGTCGCTGCTCAGCTCTCTATGAACTCCACACCTCCACTGCTGCG 1385
60 sArgValAlaGluLeuAsnValGlnValGlySerProThrHisPro.....91
1386 G.....ATCACACTCGCTGCTTCCACAAATGAC...ACCC 1420

```

```

77 yrLeuLeuProGlnAlaIleGlnValGlySerProThrHisPro.....91
1421 ACTTGCTGGCGGGTCTCCGAGTATGAGTGCAGAGGACCCCAAGTGG 1470
92 SerTyrProSer.....GlyHisAlaThrGlnAsnGlyAl 103
1471 GAGTATCCAGGAGAAAGCTCAGCTGGGAGGCCCTCGGCGAAGGTG 1520
103 aPheAlaThrValLeuLysAla.....LeuIleGlyLeuAspArgGlyGly 119
1521 CTTGGCGAGGTGGTGGTGGGAGCGGTGGCATCGACAAAGACCGGC 1570
119 LysPheProAsnProVal.....PheProSerAspAspGlyLeu 132
1571 CCAAGATGACGAGCGACCTGGCAGTGAAGATGCTGAAGACGATGCAAC 1620
133 Glu..... 133
1621 GAGAAGGATCTTTCTGATCTGCTGCTGAGATGGAATGATGAAGATGAT 1670
134 .....LeuIleAsnPheGluGlyAlaCysLeuThrTyrG 145
1671 TGGGAGCATATAAATAATATCATCAATCTCTAGGAGCGTGC...ACCCAAG 1717
145 LysGlyGluIleAsnLysLeuAlaValAlaPheGly..... 158
1718 ATGGCCCACTCTACGTGATAGTGCATATGCTCCCAAGGGAACCTGCGT 1767
159 .....ArgGlnMetLeuGlyTlleHisTyrArgPheAspG 170
1768 GATACTTGGCGACCGCGCGCCACCTGGCATGAGTATCTCTTTGAC... 1815
170 YlleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisG 187
1816 .....ATCAACAGAAATCTCTG 1831
187 LngLufLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPhe 203
1832 AAGACGACATGACCTTCAAGGACCTAGTGTCTTGACAGTACCAACTGGCC 1881
204 ThrGly 205
1882 AGGGGA 1887
seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-885-418-5
seq_documentation_block:
Sequence 5: Application US/08885418
Patent No. 5925528
GENERAL INFORMATION:
APPLICANT: Chiu, Ing-Ming
APPLICANT: Poulin, Matthew L
TITLE OF INVENTION: Acidic Fibroblast Growth Factor (aFGF)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ing-Ming Chiu
STREET: S2052 Davis Medical Research Center, 480 West
STREET: 9th Avenue
CITY: Columbus
STATE: Ohio
COUNTRY: USA
ZIP: 43210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/885,418
FILING DATE:
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614)-293-8093

```


LOCATION: 325..2517
FEATURE: 3'UTR
NAME/KEY: 2518..2681
LOCATION: 2518..2681
PUBLICATION INFORMATION:
AUTHORS: Poulin, Matthew L
AUTHORS: Chiu, Ing-Ming
TITLE: Nucleotide sequences of two new
TITLE: (No. 592528)phthalum viridescens) fibroblast growth
TITLE: factor receptor-2 variants
JOURNAL: Biochim. Biophys. Acta
VOLUME: 1220
PAGES: 209-211
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 9: FROM 1 TO 2681
US-08-885-418-9

alignment_scores:
Quality: 75.00 Length: 202
Ratio: 0.833 Gaps: 10
Percent Similarity: 44.554 Percent Identity: 22.277

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-885-418-9

Align seg 1/1 to: US-08-885-418-9 from: 1 to: 2681

44 LeuAspAlaSpPheAspLeuLeuGluAsnAspGluLeuLeu 60
1342 GTGCTGCTGACTCCAGCTTCTATGAACCTCCACACCTCCTGTCGG 1391
60 sargValAlaGluLeuAsnAlaGlnAsnProAsnGlnValThr 77
1392 G.....ATCACCACCTCGCTGCTCTCCAAACATGAC...ACCC 1426
77 yLeuLeuProGlnAlaLeuGlnValGlySerProThrHisPro..... 91
1427 ACTTGTGCGGGGGTCTCCGAGTATGAGTCCAGAGACCCCAAGTGG 1476
92 SerTyroProser.....GlyHisAlaThrGlnAsnGlyAla 103
1477 GAGTATCCAGGGAAGCTACGCTGGGGAAGCCCTGGCGGAGCTG 1526
103 pheAlaThrValLeuLysAla...LeuileGlyLeuAspArgGlyGly 119
1527 CTTCGGCAGGTGGTATGCGAGAGCGGTGGGATCGACAAAGCCGCG 1576
119 luCysPheProAsnProVal.....PheProSerAspAspGlyLeu 132
1577 CCAAGATGCGAGCGTGGCGAGTGAAGATGCTGAAGAGCATGCAACC 1626
133 Glu..... 133
1627 GAGAAGGATCTTTCTATCTGCTGCTGAGATGGAATGATGAAGATGAT 1676
134LeuileAsnPheGluGlyAlaCysLeuThrTyrg 145
1677 TGGGAAGCATAAAAATATCATCAATCTCTAGGAGCGTGC...ACCCAAG 1723
145 luGlyGluLeuAsnLysLeuAlaValAsnValAlaPheGly..... 158
1724 ATGGCCCATCTAGCTGATGATGATGATGCTCCCAAGGGGACTTGGCT 1773
159ArgGlnMetLeuGlyIleHisTyrgPheAspGly 170
1774 GAATACTTGGCCACCCCGCCCGCCCTGGCATGGAGTACTCTCTGAC... 1821
170 yIleGlnLeuLeuLeuGlyGluThrIleThrValargThrLeuHis 187
1822ATCAACAGAAATTCCTG 1837
187 InGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPhe 203

1838 AAGAGCAGATGACCTTCAAGGACCTAGTGTCTTGACACGTACCAACTGGCC 1887
204 ThrGly 205
1888 AGGGGA 1893

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq.US-08-619-198-2

seq_documentation_block:

Sequence 2, Application US/08619198

Patent No. 5885831

GENERAL INFORMATION:

APPLICANT: Young, Michael W.

APPLICANT: Sehgal, Amita

APPLICANT: Voshall, Leslie B.

APPLICANT: Price, Jeffrey L.

APPLICANT: Myers, Michael

TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED

TITLE OF INVENTION: WITH CIRCADIAN RHYTHMS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS: 11

ADDRESS: Klauber & Jackson

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/619.198

FILING DATE: 20-MAR-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-128A CPl

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201.487-5800

TELEFAX: 201.343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3369 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

DESCRIPTION: cDNA sequence wherein R at position 1335

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Drosophila melanogaster

FEATURE:

NAME/KEY: CDS

LOCATION: 1..3369

US-08-619-198-2

alignment_scores:

Quality: 74.00 Length: 141

Ratio: 1.042 Gaps: 7

Percent Similarity: 50.355 Percent Identity: 28.369

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-619-198-2/rev.

Align seg 1/1 to reverse of: US-08-619-198-2 from: 1 to: 3369

76 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProse 92
 2071 ACGTATCAATGCCCTGGA.....GGACTCCCGGATGCCGTC 2034
 92 rTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuL 109
 2033 ACCACCAAGT.....GCATTCGCC..... 2015
 109 ysaAlaLeuLeuGlyLeuAspArg...GlyGlyGlyCysPhePro..... 122
 2014TTAGTAGGGCTTCAGGTACCTGCTCCCTGTCGGG 1979
 123 AsnProValPheProSerAspGlyLeuGluLeuLeuAsnPheGluG 139
 1978 CATTCAGTTCAGTTCGTCACATAGGACACACCTCATAGTCAAGTAG 1929
 139 lyaAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnVal 155
 1928 CTCAAACATCGTAGTGAGATAGTGTGCTGCTCATATCCAGCTC 1879
 156 AlaPheGlyAlaGlnMetLeuGlyIleHisTyrArgPheAspGlyIleG 172
 1878 CAGTGGGGGCAAACTTACGAGTAGGTACAGCCAAAGAGTGGG 1829
 172 n.....GlyLeuLeuGlyGluThrIleThrValArgThrL 185
 1828 ATGTGTCATCGGCACCTTCGCAITGGAAAGCAGTTCATGTGCGATCC 1779
 185 euHisGlnGluMetThrPheAlaGluGluAlaThrPheGluPheArg 201
 1778 TCACCAACATCGTGTACCTTGGAGGAGAAATCCACGGTAATTCCTT 1729
 202 LeupheThrGlyGluValIle 208
 1728 CAGCAGATTGGAGATGTCATC 1708

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-198-4

seq_documentation_block:
 ; Sequence 4, Application US/08619198
 ; Patent No. 5885831
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Michael W.
 ; APPLICANT: Sehgal, Amita
 ; APPLICANT: Voshall, Leslie B.
 ; APPLICANT: Price, Jeffrey L.
 ; APPLICANT: Myers, Michael
 ; TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED
 ; WITH CIRCADIAN RHYTHMS
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/619.198
 ; FILING DATE: 20-MAR-1996
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-128A CPI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201 487-5800
 ; TELEFAX: 201 343-1684

TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4170 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; DESCRIPTION: tm cDNA sequence wherein R at position 1335
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..4170
 ; US-08-619-198-4

alignment_scores:
 Quality: 74.00 Length: 141
 Ratio: 1.042 Gaps: 7
 Percent Similarity: 50.355 Percent Identity: 28.369

alignment_block:
 US-09-151-189-2_COPY_441_676 x US-08-619-198-4/rev

Align seg 1/1 to reverse of: US-08-619-198-4 from: 1 to: 4170

76 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProse 92
 2071 ACGTATCAATGCCCTGGA.....GGAACTCCCGGATGCCGTC 2034
 92 rTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuL 109
 2033 ACCACCAAGT.....GCATTCGCC..... 2015
 109 ysaAlaLeuLeuGlyLeuAspArg...GlyGlyGlyCysPhePro..... 122
 2014TTAGTAGGGCTTCAGGTACCTGCTCCCTGTCGGG 1979
 123 AsnProValPheProSerAspGlyLeuGluLeuLeuAsnPheGluG 139
 1978 CATTCAGTTCAGTTCGTCACATAGGACACACCTCATAGTCAAGTAG 1929
 139 lyaAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnVal 155
 1928 CTCAAACATCGTAGTGAGATAGTGTGCTGCTCATATCCAGCTC 1879
 156 AlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleG 172
 1878 CAGTGGGGGCAAACTTACGAGTAGGTACAGCCAAAGAGTGGG 1829
 172 n.....GlyLeuLeuGlyGluThrIleThrValArgThrL 185
 1828 ATGTGTCATCGGCACCTTCGCAITGGAAAGCAGTTCATGTGCGATCC 1779
 185 euHisGlnGluMetThrPheAlaGluGluAlaThrPheGluPheArg 201
 1778 TCACCAACATCGTGTACCTTGGAGGAGAAATCCACGGTAATTCCTT 1729
 202 LeupheThrGlyGluValIle 208
 1728 CAGCAGATTGGAGATGTCATC 1708

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-399-561-1

seq_documentation_block:
 ; Sequence 1, Application US/08399561
 ; Patent No. 5792903
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirschberg, Joseph
 ; APPLICANT: Cunningham Jr., Francis X.

APPLICANT: Gantt, Elisabeth
TITLE OF INVENTION: Lycopene Cyclase Gene
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 5792903thwestern Highway, Suite 410
CITY: Farmington Hills
STATE: MI
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,561
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-305 (Hebrew Univ.)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4928 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2029..3261
US-08-399-561-1

alignment_scores:

Quality:	74.00	Length:	162
Ratio:	0.892	Gaps:	12
Percent Similarity:	51.235	Percent Identity:	29.630

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-399-561-1/rev ...

Align seg 1/1 to reverse of: US-08-399-561-1 from: 1 to: 4928

```
28 ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrLeAlaGlyAspLe 44
|||||:  ::::|  |||  |||:  :::  ::
866 CGCCCTGCG...ATTGGTGGG...TTGCCAACACCGCTTAAGCCCGCAT 823
44 uasAlaaspPheaspilesrLeuLeuGluAsnAspGluLeuLeuLysA 61
|||||:  ::::|  |||  |||:  :::  ::
822 TAGTACTGACCTCGAT.....GACGGCGTGAAGCCCTCGGCA 785
61 tgvAlaAlaGluIleAsnAlaAlaGlnAsnProAsnGluValThrTyr 77
::  |||:  ::::|  |||  |||:  :::  ::
784 AA...GCAGACTAGAGCGGA..... 765
78 LeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPr 94
|||||:  ::::|  |||  |||:  :::  ::
764 CTGATTCAGACGACATCGAIGTCACCATGCCGCGGCGAG...TATCGTCC 718
94 oSerGlyHisAla...ThrGlnAsnGlyAlaPheAlaThrValLeuLysA 110
|||||:  ::::|  |||  |||:  :::  ::
717 CCAGGACACACCTTCACCCGTCGAAGCAGCATCGCATTCGCGACA 668
110 laLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProValPhe 126
|||||:  ::::|  |||  |||:  :::  ::
667 TCCTGCTGGGCGTT.....GGCTATCAAGTCGCTGAGGGGCCAGAAATG 624
```



```

153 TCTTGAGCTCTACGCGAGAACTTGTACAAGCTGGACGGCAAGCTCTGC 202
      :::::::::::::::
224 SerGlyLeuValThrGlyValAlaAspCys 234
      :::::::::::::::
203 GAGGG...GCCTTACCGCGCAGCACTTCTGT 232
      :::::::::::::::

seq_name: gb_est35:AL045228
seq_documentation_block:
LOCUS AL045228 468 bp mRNA EST 29-SEP-1999
DEFINITION DKF2p434H2050.r1.434 (synonym: htes3) Homo sapiens CDNA clone
AUTHORS Poustka,A., Klein,M., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Poustka, et al.)
JOURNAL Unpublished (1999)
COMMENT On May 1, 1997 this sequence version replaced gi:2059563.
          Contact: Poustka A.J.
          Department Lehrach
          Max-Planck-Institute for Molecular Genetics
          Ihnestrasse 73, 14195 Berlin, Germany
          Tel: +49-30-84131623
          Fax: +49-30-84131128
          Email: pousta@mpg-berlin-dahlem.mpg.de
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
Sequenced by DKFZ within the cDNA sequencing consortium of the
German Genome Project.
No SI sequence available.
This clone is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
    source
        1..468
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKF2p434H2050"
            /clone_lib="434 (synonym: htes3)"
            /tissue_type="testis"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT    98 a 137 c 156 g
ORIGIN
    alignment_scores:
        Quality: 86.50 Length: 149
        Ratio: 1.153 Gaps: 9
        Percent Similarity: 50.336 Percent Identity: 28.859

alignment_block:
US-09-151-189-2_COPY_441_676 x AL045228
Align seg 1/1 to: AL045228 from: 1 to: 468
1 HistYrPheArgLeuGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
      :::::::::::::::
29 CACTACACAGCTC.....TCTGCAGAGCAGCTCGTCGGAG..... 67
      :::::::::::::::
17 stycGlnys.....TgcGlnValHisArgPheAlaArgProGluAla 32
      :::::::::::::::
68 .....CAGAGGCACTGTGGAGATCCAGG..... 94
      :::::::::::::::

```

```

32 euGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 48
      :::::::::::::::
95 .....CACCAGCTGGAGAGTGCACGGCTTGT..... 121
      :::::::::::::::
49 AspileSerLeuLeuGluAsnAsp.....GluLeuLeuLys 61
      :::::::::::::::
122 .....TTTCTCTTAGAGATGAGAAACACATTTCAGGAGATGCTGAAGC 165
      :::::::::::::::
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnGlnValThrTyrL 78
      :::::::::::::::
166 AGTGTCTGAGGCTCACCACCCCGAGCGCCAGACGCTCTCTTGAGG 215
      :::::::::::::::
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 94
      :::::::::::::::
216 TGCACCCCGAGTCACGCTCTCCGGC...CCTGAGCACCCAGAG..... 256
      :::::::::::::::
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAla 111
      :::::::::::::::
257 GGAGCCCAAGCGTGTATCTGGTCTGCAGAGCAACACTCGCCTCCCTG 306
      :::::::::::::::
111 ulleGlyLeuAspArgGlyGlyCysPheProAsnProValPhePro. 127
      :::::::::::::::
307 GGATGGCTGGAACAGCCAGCGCTCTGACCCACACGCCCTTAAAGCCTC 356
      :::::::::::::::
128 .....SerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 140
      :::::::::::::::
357 TAGCAGTGGGGGCTGGTGCAGGGGCTGCAGAGGTGGAGGGGGCC 403
      :::::::::::::::

seq_name: gb_est47:AW401290
seq_documentation_block:
LOCUS AW401290 289 bp mRNA EST 07-FEB-2000
DEFINITION Lamdigest491est.L.digitata gametophyte Lambda ZapII Laminaria
digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.
ACCESSION AW401290
VERSION AW401290.1 GI:6919794
KEYWORDS EST.
SOURCE Laminaria digitata
ORGANISM Laminaria digitata
Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
Phaeophyceae; Laminariales; Laminariaceae; Laminaria.
1 (bases 1 to 289)
Crepineau,F., Roscoe,T., Kaas,R., Kloareg,B. and Boyen,C.
Characterisation of complementary DNAs from the Expressed Sequence
Tag analysis of life cycle stages of Laminaria digitata
(Phaeophyceae)
Unpublished (2000)
On Dec 20, 1995 this sequence version replaced gi:1135359.
Other ESTs: Lamdigest4917est
Contact: Boyen C
Centre d'Etudes Oceanologique et de Biologie Marine
CNRS-LRP9042, Universite P. & M. Curie
BP74, F-29682 Roscoff cedex, France
Tel: 33 2 98 29 23 32
Fax: 33 2 98 29 23 24
Email: est@sb-roscoff.fr.

FEATURES
    source
        1..289
            /organism="Laminaria digitata"
            /db_xref="taxon:80365"
            /clone_lib="L.digitata gametophyte Lambda ZapII"
            /dev_stage="gametophyte"
            /lab_host="SOLR"
            /note="vector: pBluescript SK+; Site_1: EcoRI; Site_2:
            XhoI; The Laminaria gametophyte library, constructed by T
            Roscoe and F Crepineau, was oligo-(dT) primed and
            directionally cloned into a Uni-ZAPTM XR vector
            (Stratagene, la Jolla, CA, USA) using total mRNA from
            gametophytes provided by R Kaas, IFREMER Nantes, France."

BASE COUNT    71 a 88 c 77 g 53 t
ORIGIN

```


53 euGluAnaspGluLeuLeuLysArgVal...AlaGluLeuLeuAlaAla 68
 333 TCCAGTCCACCAGATAGCTAGATAGAGGTGTGATTGTGTGATCC 304
 69 GlnAsnPro.....AsnAsnGluValThrThrLeuLeu.....ProG1 81
 303 ACAAACCTGAGCTAGACACAGAGTGTGGGCGCATACACAACTCC 254
 81 nAlaLeuGluValGlySerProThrHisProSerTyrProSerGlyHisA 98
 253 AGCC.....AGACATAAAGTGTCTCCAGTCCCTCCACT 222
 98 laThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuLeuGlyLeu 114
 221 TGACTAGAGGCCACTGTTTCCCTAGTGGATCCCTGTGGGGCTG 172
 115 AspArgGlyGlyGluCysPheProAsnProVal 125
 171 CCAGCGGGGACCCCTGCACTCCTCAGCCCTTG 139
 seq_name: gb_est38:AL120080

seq_documentation_block: 750 bp mRNA EST 27-SEP-1999
 LOCUS AL120080
 DEFINITION DKF2p761M212.r1.761 (synonym: hamy2) Homo sapiens CDNA clone
 VERSION AL120080
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 OTTENWAEELDER, B., OBERMAIER, B., MEVES, H.W., GASSENHUBER, J. and
 WIEMANN, S.
 EST (Ottenwaeelder, et al.)
 Unpublished (1999)
 On Jun 22, 1998 this sequence version replaced gi:3247433.
 Contact: Ottenwaeelder B
 MIPS
 Am Klopferspitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by MediGenomix within the CDNA sequencing consortium of
 the German Genome project.
 No 1 sequence available.
 This clone is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1..750
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKF2p761M212"
 /clone_lib="761 (synonym: hamy2)"
 /tissue_type="amygdala"
 /dev_stage="adult"
 /lab_host="DB108"
 /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 173 a 181 c 213 g 183 t
 ORIGIN

alignment_scores:
 Quality: 78.50 Length: 135
 Ratio: 1.172 Gaps: 7
 Percent Similarity: 49.630 Percent Identity: 28.148

alignment_block:
 US-09-151-189-2_COPY_441_676 x AL120080/rev

Align seg 1/1 to reverse of: AL120080 from: 1 to: 750

89 ThrHisProSer...TyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 104
 386 ACACATGAGCCCTAATACAGGTGGGGGAGAGAGCTCAAGAAAAAGGTG 337
 104 eAlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGlyGluCysP 121
 336 GGAGGGGATCTTTTACTCACTCTCTGCTGTGAT..... 303
 121 heProAsnProValPheProSerAspArgGlyLeuGluLeuLeuAsnPhe 137
 302ACCCAGCGACCTCCCTTGTCTGCTGCT 276
 138 GluGlyAlaCysLeuThrTyrGluGlyLeuLeuAsnLysLeuAlaValAs 154
 275 GAGGTGCTGTC.....GCCCTGCCGTAGC 250
 154 nValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyI 171
 249 CATAGCCTTTGGGCCCATATCTTTCGCGTA.....GCAGGAC 212
 171 leGlnGlyLeuLeu.....GlyGluThrIleThrValArgThr 184
 211 TTGCAGTAATATCTCTCACCATGCAGCGCCACAGTGTCTGTCAGATT 162
 185 Leu.....HisGlnGluLeuMetThrPheAlaGluLeuAlaThrPh 198
 161 CTCTTCGACACCATGCACAGGAGGATTATGAGACGCTGTGCTT 112
 198 eGluPheArgLeuPheThrGlyGluValLeuLysLeuPheGlnAspGlyT 215
 111 CGCCTGAACCTCTTC...GCCAAGTAACCGCTCTTCTGCACACACCCCA 65
 215 hrPhe 216
 64 CATT 60
 seq_name: gb_est24:AI232282

seq_documentation_block: 618 bp mRNA EST 31-JAN-1999
 LOCUS AI232282
 DEFINITION EST228970 Normalized rat kidney, Bento Soares Rattus sp. CDNA clone
 RKBZ71 3' end, mRNA sequence.
 ACCESSION AI232282
 VERSION AI232282.1 GI:3816162
 KEYWORDS EST
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Jan 14, 1998 this sequence version replaced gi:1797454.
 Other ESTs: TC62480
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research,
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..618
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2037931"
 /db_xref="taxon:10118"

FEATURES
 source

/clone="RKIB271"
/clone_lib="Normalized rat kidney, Bento Soares"
/note="Organ: kidney; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 132 a 127 c 187 g 172 t
ORIGIN

alignment_scores:
Quality: 78.00 Length: 121
Ratio: 1.130 Gaps: 9
Percent Similarity: 57.025 Percent Identity: 28.099

alignment_block:
US-09-151-189-2_COPY_441_676 x AI232282/rev

Align seg 1/1 to reverse of: AI232282 from: 1 to: 618

116 ArgGlyGlyGluCysPheProAsnProValPheProSerAspGlyLe 132
|||||
614 CGAGGGGAGAGCTGTCTCCAGGAGCGCTCACCATGTCGGAGAAA...TT 568
132 uGluLeuLeuAsnPhe.....GluGlyAlaCysLeuThrTyrGluG 146
|||||
567 CGAGGTCAAACTGACATGAATCAGGAGTCCCTGAAGATTAAAG 518
146 ly.....GluLeuAsnLysLeuAlaValAsnValAlaPheGly 158
|||||
517 GCAAGATCCACATGATGTATACAGTTTACCATTATCTGGCGCAGGG 468
159 ArgGlnMetLeuGlyIleHisTyr.....ArgPheAspGlyIleGlnG 173
|||||
467 AAGAAGACGCTGACCTGATTTAACTTACCTCGCTTCAAT..... 429
173 yLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeu 190
|||||
428GAATCCACCATTTCTGTATACACCTT.....GACGGTA 395
190 etThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThr...Gly 205
|||||
394 GCAGCTGGGACAGAGACGAGAAACCATTCCTGTTTCAGTCCAGGG 345
206 GluValIleLysLeu.....PheGlnAspGlyThrPheSerIle... 218
|||||
344 TCAGAGGTCAAGCTACCTTACCTTCCAGAGCAAGACTTCAAGGTGAC 295
219AspGly 220
294 GTCGCTGATGGA 282

seq_name: gb_est35:AI854183

seq_documentation_block:
LOCUS AI854183 498 bp mRNA EST 15-JUL-1999
DEFINITION UI-M-BH0-ajz-c-01-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone
UI-M-BH0-ajz-c-01-0-UI 3', mRNA sequence.

ACCESSION AI854183
VERSION AI854183.1 GI:5498089
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

COMMENT
On May 11, 1999 this sequence version replaced gi:4776633.
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA
Tel.: 301 443 1706
Fax: 301 443 9890
Email: mstremail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized cerebellum library cDNA library Preparation: M.B. Soares
Lab Clone distribution: NIH BMAP cDNA clones will be made available
by the means that is soon to be determined. When NIH determines the
means for distribution of the BMAP cDNA clones, this record will be
updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA-yes.

FEATURES
source
1. 498
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH0-ajz-c-01-0-UI"
/clone_lib="NIH_BMAP_M_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73P-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The driver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of
these ten regions of the mouse brain.
TAG_LIB=NIH_BMAP_M_S1
TAG_TISSUE=cerebellum
TAG_SEQ=GACTC"

BASE COUNT 104 a 164 c 119 g 110 t 1 others
ORIGIN

alignment_scores:
Quality: 77.50 Length: 130
Ratio: 1.123 Gaps: 9
Percent Similarity: 53.077 Percent Identity: 27.692

alignment_block:
US-09-151-189-2_COPY_441_676 x AI854183

Align seg 1/1 to: AI854183 from: 1 to: 498

61 ArgValAlaGluLeuAsnAlaAlaGlnAsnProAsnAsnGlu.ValThrT 77
|||||
59 AAATTAATCATCTTCATGTTTCATCTCAACCAATCCAGATCCCTCATG 108
77 yrlLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyr 93
|||||
109 ACACACTGCCGGGA.....AACAGCCCTGCCCTGCCCTCCCAT 146
94 ProSerGly.....HisAlaThrGlnAsnGlyAlaPheAl 105
|||||
147 GGTTCAGGTGGGAGACACCATTCACAGCAGGAAGTCTGTCGCCCAT 196
105 aThrValLeuLysAla.....LeuIleGly.... 113
|||||
197 GCGTGTCTGCAGCAGGAGCGCTCCACCCACCGCTCTTCCTAGGAGCCT 246
114LeuAspArgGlyGlyGluCysPheProAsnProVal 125
|||||
247 CCCATCATGCTCTTCTTGGCGCGCGGCGGAGCAGCTCCG..... 287
126 PheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLe 142

```

288 .....CAGGCGCTCGAAGTCCGGAACGACAG.....GTCT 319
142 uthrTyrglu.....GlygluileasnLysLeuAlaVala 154
    ||||| ..... |||||
320 TACTCAGGCGGCTCTGGGGCGTGGAGCGCTGCACGCACTTGCCTAA 369
154 sn.....ValAlaPheGlyArgGlnMetLeuGly 163
    :: ..... |||||
370 GCCGAGGCGCTCAGCTCCACAGCGCGCCAGATGTCGGG 407

seq_name: gb_est25:AI304074

seq_documentation_block:
LOCUS AI304074 803 bp mRNA EST 08-DEC-1998
DEFINITION ui63h06.v1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:1887131.5' similar to TR:035678 035678 MONOGLYCERIDE LIPASE.
; mRNA sequence.
ACCESSION AI304074
VERSION AI304074.1 GI:3987824
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 803)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:971455
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 422.
Location/Qualifiers
1..803
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1887131"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DR10B"
/notes="Organ: liver; Vector: pME18S-FL3; Site: 1: DraIII
(CACTGTGTG); Site: 2: DraIII (CACCAGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCAGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
primer CGACCTGCGCTCGAGCACA."

```

```

167 a 215 c 213 g 207 t 1 others
ORIGIN

```

```

alignment_scores:
Quality: 77.00 Length: 88
Ratio: 1.453 Gaps: 2
Percent Similarity: 60.227 Percent Identity: 28.409

alignment_block:
US-09-151-189-2_COPY_441_676 x AI304074/rev ..
Align seg 1/1 to reverse of: AI304074 from: 1 to: 803

34 GlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAsp11 50
||||| ||||| ..... |||||
272 GGCACGCGCCATGGA...GTGGCCGAGGAGGAGAGTGGGACGCTCGGGCT 226

50 eSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnA 67
||||| ||||| ..... |||||
225 AGTCCTTCTGGTGGTGTCCAGCTGTTCAGCACATCTCTGACAAAAACT 176

67 IaLaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 83
||||| ..... |||||
175 TGGAAAGTCCGACACCACTCTCTCCCTCACTCTGCCCATGGCCAAAC 126

84 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrG1 100
||||| ..... |||||
125 ATGSGTCATGGGCAAAATACACAGCATGTCCAGCCCTTCAACATATGAGCCA 76

100 nasngIyAlaPheAlaThrValLeuLysAlaLeuIleGly...LeuAspA 116
||||| ..... |||||
75 GCTCATCATACGCGCCACAGTGAAGAAACAGCTCCATGGGACACAAAGATG 26

116 rGlyGlyGluCys 120
||||| ..... |||||
25 AGGCGCTTGTAGTGT 12

```

FEATURES

```

source
1..803
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1887131"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DR10B"
/notes="Organ: liver; Vector: pME18S-FL3; Site: 1: DraIII
(CACTGTGTG); Site: 2: DraIII (CACCAGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCAGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
primer CGACCTGCGCTCGAGCACA."

```

```

BASE COUNT 167 a 215 c 213 g 207 t 1 others
ORIGIN

```
